(51) International Patent Classification 6 :		(11) International Publication Number:	WO 95/15392
C12N 15/82, 15/60, 15/11, A01H 5/00, 5/10	A1	(43) International Publication Date:	8 June 1995 (08.06.95)

(21) International Application Number: PCT/US94/13190

(22) International Filing Date: 21 November 1994 (21.11.94)

(30) Priority Data:
08/160,117   30 November 1993 (30.11.93)   US
08/261,661   17 June 1994 (17.06.94)   US

(60) Parent Applications or Grants
  (63) Related by Continuation
  US   08/160,117 (CIP)
  Filed on   30 November 1993 (30.11.93)
  US   08/261,661 (CIP)
  Filed on   17 June 1994 (17.06.94)

(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): FALCO, Saverio, Carl [US/US]; 1902 Millers Road, Arden, DE 19810 (US). KEELER, Sharon, Jo [US/US]; 1320 Barksdale Road, Newark, DE 19711-2636 (US). RICE, Janet, Ann [US/US]; 126 Median Drive, Wilmington, DE 19803 (US).

(74) Agents: SIEGELL, Barbara, C. et al.; E.I. du Pont de Nemours and Company, Legal/Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AU, BR, CA, HU, JP, KR, PL, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
  With international search report.
  Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE CONTENT OF THE SEEDS OF CORN, SOYBEAN AND RAPESEED PLANTS

(57) Abstract

This invention relates to three chimeric genes, the first encoding dihydrodipicolinic acid synthase (DHDPS), which is insensitive to inhibition by lysine and operably linked to a plant chloroplast transit sequence, a second encoding a lysine-rich protein, and a third encoding a plant lysine ketoglutarate reductase, all operably linked to plant seed-specific regulatory sequences. Methods for their use to produce increased levels of lysine in the seeds of transformed plants are provided. Also provided are transformed corn, rapeseed and soybean plants wherein the seeds accumulate lysine to higher levels than untransformed plants.

1

TITLE

CHIMERIC GENES AND METHODS FOR
INCREASING THE LYSINE CONTENT OF THE
SEEDS OF CORN, SOYBEAN AND RAPESEED PLANTS

TECHNICAL FIELD

This invention relates to three chimeric genes, the first encoding dihydrodipicolinic acid synthase (DHDPS), which is insensitive to inhibition by lysine and operably linked to a plant chloroplast transit sequence, a second encoding a lysine-rich protein, and a third encoding a plant lysine ketoglutarate reductase, all operably linked to plant seed-specific regulatory sequences. Methods for their use to produce increased levels of lysine in the seeds of transformed plants are provided. Also provided are transformed corn, rapeseed and soybean plants wherein the seeds accumulate lysine to higher levels than untransformed plants.

BACKGROUND OF THE INVENTION

Human food and animal feed derived from many grains are deficient in some of the ten essential amino acids which are required in the animal diet. In corn (Zea mays L.), lysine is the most limiting amino acid for the dietary requirements of many animals. Meal derived from other crop plants, e.g., soybean (Glycine max L.) or Canola (Brassica napus), is used as an additive to corn based animal feeds to supplement this lysine deficiency. Also, additional lysine, produced via fermentation of microbes, is used as a supplement in animal feeds. An increase in the lysine content of meal derived from plant sources would reduce or eliminate the need to supplement mixed grain feeds with microbially produced lysine.

The amino acid content of seeds is determined primarily (90-99%) by the amino acid composition of the proteins in the seed and to a lesser extent (1-10%) by

2

the free amino acid pools. The quantity of total
protein in seeds varies from about 10% of the dry weight
in cereals to 20-40% of the dry weight of legumes. Much
of the protein-bound amino acids is contained in the
5   seed storage proteins which are synthesized during seed
development and which serve as a major nutrient reserve
following germination. In many seeds the storage
proteins account for 50% or more of the total protein.
        To improve the amino acid composition of seeds
10   genetic engineering technology is being used to isolate,
and express genes for storage proteins in transgenic
plants. For example, a gene from Brazil nut for a seed
2S albumin composed of 26% sulfur-containing amino acids
has been isolated [Altenbach et al. (1987) Plant Mol.
15   Biol. 8:239-250] and expressed in the seeds of
transformed tobacco under the control of the regulatory
sequences from a bean phaseolin storage protein gene.
The accumulation of the sulfur-rich protein in the
tobacco seeds resulted in an up to 30% increase in the
20   level of methionine in the seeds [Altenbach et al.
(1989) Plant Mol. Biol. 13:513-522]. However, no plant
seed storage proteins similarly enriched in lysine
relative to average lysine content of plant proteins
have been identified to date, preventing this approach
25   from being used to increase lysine.
        An alternative approach is to increase the
production and accumulation of lysine via genetic
engineering technology. Lysine, along with threonine,
methionine and isoleucine, are amino acids derived from
30   aspartate, and regulation of the biosynthesis of each
member of this family is complex, interconnected, and
not well understood, especially in plants. Regulation
of the metabolic flow in the pathway appears to be
primarily via end products in plants. The aspartate
35   family pathway is also regulated at the branch-point

reactions. For lysine this is the condensation of
aspartyl β-semialdehyde with pyruvate catalyzed by
dihydrodipicolinic acid synthase (DHDPS).

The E. coli dapA gene encodes a DHDPS enzyme that
is about 20-fold less sensitive to inhibition by lysine
than than a typical plant DHDPS enzyme, e.g., wheat germ
DHDPS. The E. coli dapA gene has been linked to the 35S
promoter of Cauliflower Mosaic Virus and a plant
chloroplast transit sequence. The chimeric gene was
introduced into tobacco cells via transformation and
shown to cause a substantial increase in free lysine
levels in leaves [Glassman et al. (1989) PCT Patent
Appl. PCT/US89/01309, Shaul et al. (1992) Plant Jour.
2:203-209, Galili et al. (1992) EPO Patent Appl.
91119328.2, Falco, PCT/US93/02480 (International
Publication Number WO 93/19190). However, the lysine
content of the seeds was not increased in any of the
transformed plants described in these studies. The same
chimeric gene was also introduced into potato cells and
lead to small increases in free lysine in leaves, roots
and tubers of regenerated plants [Galili et al. (1992)
EPO Patent Appl. 91119328.2, Perl et al. (1992) Plant
Mol. Biol. 19:815-823].

Falco, PCT/US93/02480 (International Publication
Number WO 93/19190, linked the E. coli dapA gene to the
bean phaseolin promoter and a plant chloroplast transit
sequence to increase expression in seeds, but still
observed no increase in the lysine level in seeds. As
noted above, the first step in the lysine biosynthetic
pathway is catalyzed by aspartokinase (AK), and this
enzyme has been found to be an important target for
regulation in many organisms. Falco isolated a mutant
of the E. coli lysC gene, which encoded a lysine-
feedback-insensitive AK, and linked it to the bean
phaseolin promoter and a plant chloroplast transit

4

sequence. Expression of this chimeric gene in the seeds
of transformed tobacco lead to a substantial increase in
the level of threonine, but not lysine. Galili et al.
(1992) EPO Patent Appl. 91119328.2 suggest that
5   transforming plants with chimeric genes linking seed-
specific promoters to a plant chloroplast transit
sequence/E. coli dapA gene and plant chloroplast transit
sequence/mutant E. coli lysC gene will lead to increased
lysine levels in seeds. Falco, PCT/US93/02480
10 (International Publication Number WO 93/19190) carried
out this experiment by transforming tobacco with a
construct containing both the chimeric genes, bean
phaseolin promoter/plant chloroplast transit
sequence/E. coli dapA gene and bean phaseolin
15 promoter/plant chloroplast transit sequence/mutant
E. coli lysC gene. Simultaneous expression of both
genes had no significant effect on the lysine content of
the seeds. However, it was noted that a breakdown
product of lysine, α-amino adipic acid, built up in the
20 seeds. This suggested that the accumulation of free
lysine in seeds was prevented because of lysine
catabolism. In an effort to increase the rate of
biosynthesis of lysine, Falco, PCT/US93/02480
(International Publication Number WO 93/19190, isolated
25 the Corynebacterium glutamicum dapA gene which encodes a
completely lysine insensitive DHDPS enzyme. Falco
transformed tobacco with a construct containing the
chimeric gene, bean phaseolin promoter/plant chloroplast
transit sequence/Corynebacterium glutamicum dapA gene
30 linked to bean phaseolin promoter/plant chloroplast
transit sequence/mutant E. coli lysC gene. Simultaneous
expression of both these lysine-insensitve enzymes still
had no significant effect on the lysine content of the
seeds.

5

Thus, it is clear that the limited understanding of the details of the regulation of the lysine biosynthetic pathway in plants, particularly in seeds, makes the application of genetic engineering technology to

5   increase lysine content uncertain. It is not known, for most plants, whether lysine is synthesized in seeds or transported to the seeds from leaves. In addition, little is known about storage or catabolism of lysine in seeds. Because free amino acids make up only a small

10 fraction of the total amino acid content of seeds, over-accumulation must be many-fold in order to significantly affect the total amino acid composition of the seeds. In addition, the effects of over-accumulation of a free amino acid such as lysine on seed development and

15 viability is not known.

No method to increase the lysine content of seeds via genetic engineering and no examples of seeds having increased lysine levels obtained via genetic engineering were known before the invention described herein.

20   SUMMARY OF THE INVENTION

This invention concerns a novel chimeric gene, and plants transformed using said novel gene, wherein a nucleic acid fragment encoding dihydrodipicolinic acid synthase, which is insensitive to inhibition by lysine,

25 is operably linked to a plant chloroplast transit sequence and to a plant seed-specific regulatory sequence. In a preferred embodiment, the nucleic acid fragment encoding dihydrodipicolinic acid synthase comprises the nucleotide sequence shown in SEQ ID NO:3:

30 encoding dihydrodipicolinic acid synthase from _Corynebacterium glutamicum_. In especially preferred embodiments, the plant chloroplast transit sequence is derived from a gene encoding the small subunit of ribulose 1,5-bisphosphate carboxylase, and the seed-

35 specific regulatory sequence is from the gene encoding

the β subunit of the seed storage protein phaseolin from the bean Phaseolus vulgaris, the Kunitz trypsin inhibitor 3 gene of Glycine max, or a monocot embryo-specific promoter, preferably from the globulin 1 gene
5   from Zea maize.

The genes described may be used, for example, for transforming plants, preferably corn, rapeseed or soybean plants. Also claimed are seeds obtained from the transformed plants. The invention can produce
10   transformed plants wherein the seeds of the plants accumulate lysine to a level at least ten percent higher than in seeds of untransformed plants, preferably ten to four hundred percent higher than in untransformed plants.

15   The invention further concerns a method for obtaining a plant, preferably a corn, rapeseed or soybean plant wherein the seeds of the plants accumulate lysine to a level from ten percent to four hundred percent higher than seeds of untransformed plants
20   comprising:

(a)   transforming plant cells, preferably corn, rapeseed or soybean cells, with the chimeric gene described above;

(b)   regenerating fertile mature plants from
25   the transformed plant cells obtained from step (a) under conditions suitable to obtain seeds;

(c)   screening the progeny seed of step (b) for lysine content; and

(d)   selecting those lines whose seeds contain
30   increased levels of lysine. Transformed plants obtained from this method are also claimed.

The invention additionally concerns a nucleic acid fragment comprising

(a)   a first chimeric gene described above and

7

(b)   a second chimeric gene wherein a nucleic
acid fragment encoding a lysine-rich protein, wherein
the weight percent lysine is at least 15%, is operably
linked to a plant seed-specific regulatory sequence.

5   Also described is a nucleic acid fragment
comprising

(a)   a first chimeric gene as described above
and

(b)   a second chimeric gene wherein a nucleic
10   acid fragment encoding a lysine-rich protein comprises a
nucleic acid sequence encoding a protein comprising n
heptad units (d e f g a b c), each heptad being either
the same or different, wherein:

n is at least 4;

15   a and d are independently selected from
the group consisting of Met, Leu,
Val, Ile and Thr;

e and g are independently selected from
the group consisting of the acid/base

20   pairs Glu/Lys, Lys/Glu, Arg/Glu,
Arg/Asp, Lys/Asp, Glu/Arg, Asp/Arg
and Asp/Lys; and

b, c and f are independently any amino
acids except Gly or Pro and at least

25   two amino acids of b, c and f in each
heptad are selected from the group
consisting of Glu, Lys, Asp, Arg,
His, Thr, Ser, Asn, Ala, Gln and Cys,
said nucleic acid fragment is operably linked to a plant

30   seed-specific regulatory sequence.

Further described herein is a nucleic acid fragment
comprising

(a)   a first chimeric gene descibed above; and

(b)   a second chimeric gene wherein a nucleic
35   acid fragment encoding a lysine-rich protein comprises a

8

nucleic acid sequence encoding a protein having the
amino acid sequence (MEEKLKA)$_6$(MEEKMKA)$_2$ is operably
linked to a plant seed-specific regulatory sequence.

Also claimed herein are plants containing various
5   embodiments of the described first chimeric genes and
second chimeric genes and the described nucleic acid
fragments and seeds obtained from such plants.

The invention further concerns a nucleic acid
fragment comprising
10   (a)   a first chimeric gene as described above
and

(b)   a second chimeric gene wherein a nucleic
acid fragment encoding a lysine ketoglutarate reductase
is operably linked in the sense or antisense orientation
15   to a plant seed-specific regulatory sequence. Also
claimed is a plant comprising in its genome that nucleic
acid fragment and a seed obtained from such plant.

BRIEF DESCRIPTION OF THE
DRAWINGS AND SEQUENCE DESCRIPTIONS

20   The invention can be more fully understood from the
following detailed description and the accompanying
drawings and the sequence descriptions which form a part
of this application.

Figure 1 shows an alpha helix from the side and top
25   views.

Figure 2 shows end (Figure 2a) and side (Figure 2b)
views of an alpha helical coiled-coil structure.

Figure 3 shows the chemical structure of leucine
and methionine emphasizing their similar shapes.

30   Figure 4 shows a schematic representation of a
seed-specific gene expression cassette.

Figure 5A shows a map of the binary plasmid vector
pZS199; Figure 5B shows a map of the binary plasmid
vector pFS926.

Figure 6A shows a map of the plasmid vector pBT603;
Figure 6B shows a map of the plasmid vector pBT614.

Figure 7 depicts the strategy for creating a vector
(pSK5) for use in construction and expression of the SSP
gene sequences.

Figure 8 shows the strategy for inserting oligo-
nucleotide sequences into the unique Ear I site of the
base gene sequence.

Figure 9 shows the insertion of the base gene
oligonucleotides into the Nco I/EcoR I sites of pSK5 to
create the plasmid pSK6. This base gene sequence was
used as in Figure 8 to insert the various SSP coding
regions at the unique Ear I site to create the cloned
segments listed.

Figure 10 shows the insertion of the 63 bp
"segment" oligonucleotides used to create non-repetitive
gene sequences for use in the duplication scheme in
Figure 11.

Figure 11 (A and B) shows the strategy for
multiplying non-repetitive gene "segments" utilizing in-
frame fusions.

Figure 12 shows the vectors containing seed
specific promoter and 3' sequence cassettes. SSP
sequences were inserted into these vectors using the
Nco I and Asp718 sites.

Figure 13 shows a map of the binary plasmid vector
pZS97.

Figure 14 shows a map of the plasmid vector pML63.

Figure 15 shows a map of the plasmid vector pML102
carrying a chimeric gene wherein seed specific
regulatory sequences (from the soybean Kunitz trypsin
inhibitor 3 gene) are linked to a chloroplast transit
sequence (from the small subunit of soybean ribulose
bis-phosphate carboxylase) and the coding sequence for

lysine-insensitive dihydrodipicolinic acid synthase (the
dapA gene from Corynebacterium glutamicum).

SEQ ID NOS:1 and 2 were used in Example 1 as PCR
primers for the isolation of the Corynebacterium dapA

5   gene.

SEQ ID NO:3 shows the nucleotide and amino acid
sequence of the coding region of the wild type
Corynebacterium dapA gene, which encodes lysine-
insensitive DHDPS, described in Example 1.

10   SEQ ID NO:4 shows an oligonucleotide used in
Example 2 to create an Nco I site at the translation
start codon of the E. coli dapA gene.

SEQ ID NO:5 shows the nucleotide and amino acid
sequence of the coding region of the wild type E. coli

15   lysC gene, which encodes AKIII, described in Example 3.

SEQ ID NOS:6 and 7 were used in Example 3 to create
an Nco I site at the translation start codon of the
E. coli lysC gene.

SEQ ID NOS:8, 9, 10 and 11 were used in Example 4

20   to create a chloroplast transit sequence and link the
sequence to the E. coli lysC-M4, E. coli dapA and
Corynebacteria dapA genes.

SEQ ID NOS:12 and 13 were used in Example 4 to
create a Kpn I site immediately following the

25   translation stop codon of the E. coli dapA gene.

SEQ ID NOS:14 and 15 were used in Example 4 as PCR
primers to create a soybean chloroplast transit sequence
and link the sequence to the Corynebacterium dapA gene.

SEQ ID NOS:16-92 represent nucleic acid fragments

30   and the polypeptides they encode that are used to create
chimeric genes for lysine-rich synthetic seed storage
proteins suitable for expression in the seeds of plants.

SEQ ID NOS:93-98 were used in Example 12 to create
a corn chloroplast transit sequence.

SEQ ID NOS:99 and 100 were used in Example 12 as
PCR primers to create a corn chloroplast transit
sequence and link the sequence to the E. coli dapA gene.

5 The Sequence Descriptions contain the one letter
code for nucleotide sequence characters and the three
letter codes for amino acids as defined in conformity
with the IUPAC-IYUB standards described in Nucleic Acids
Research 13:3021-3030(1985) and in the Biochemical
Journal 219 (No. 2):345-373(1984), which are
10 incorporated by reference herein.

DETAILED DESCRIPTION OF THE INVENTION

The teachings below describe nucleic acid fragments
and procedures useful for increasing the accumulation of
lysine in the seeds of transformed plants, as compared
15 to levels of lysine in untransformed plants. In order
to increase the accumulation of free lysine in the seeds
of plants via genetic engineering, a determination was
made of which enyzmes in this pathway controlled the
pathway in the seeds of plants. In order to accomplish
20 this, genes encoding enzymes in the pathway were
isolated from bacteria. Intracellular localization
sequences and suitable regulatory sequences for
expression in the seeds of plants were linked to create
chimeric genes. The chimeric genes were then introduced
25 into plants via transformation and assessed for their
ability to elicit accumulation of the lysine in seeds.
Expression of lysine-insensitive dihydrodipicolinic acid
synthase (DHDPS), under control of a strong seed-
specific promoter, is shown to increase free lysine
30 levels 10 to 100 fold in corn, rapeseed and soybean
seeds.

It has been discovered that the full potential for
accumulation of excess free lysine in seeds is reduced
by lysine catabolism. Provided herein are two
35 alternative routes to prevent the loss of excess lysine

12

due to catabolism. In the first approach, lysine
catabolism is prevented through reduction in the
activity of the enzyme lysine ketoglutarate reductase
(LKR), which catalyzes the first step in lysine
5   breakdown. A procedure to isolate plant LKR genes is
provided. Chimeric genes for expression of antisense
LKR RNA or for cosuppression of LKR in the seeds of
plants are created. The chimeric gene is then linked to
the chimeric DHDPS gene and both are introduced into
10   plants via transformation simultaneously, or the genes
are brought together by crossing plants transformed
independently with each of the chimeric genes.
          In the second approach, excess free lysine is
incorporated into a form that is insensitive to
15   breakdown, e.g., by incorporating it into a di-, tri- or
oligopeptide, or a lysine-rich storage protein. The
design of polypeptides which can be expressed _in_ _vivo_ to
serve as lysine-rich seed storage proteins is provided.
Genes encoding the lysine-rich synthetic storage
20   proteins (SSP) are synthesized and chimeric genes
wherein the SSP genes are linked to suitable regulatory
sequences for expression in the seeds of plants are
created. The SSP chimeric gene is then linked to the
chimeric DHDPS gene and both are introduced into plants
25   via transformation simultaneously, or the genes are
brought together by crossing plants transformed
independently with each of the chimeric genes.
          A method for transforming plants, preferably corn,
rapeseed and soybean plants is taught herein wherein the
30   resulting seeds of the plants have at least ten percent,
preferably ten percent to 400 percent greater lysine
than the seeds of untransformed plants. Provided as
examples herein are transformed rapeseed plants with
seed lysine levels increased by 100% over untransformed
35   plants, soybean plants with seed lysine levels increased

by 400% over untransformed plants, and transformed corn
plants with seed lysine levels increased by 130% over
untransformed plants.

In the context of this disclosure, a number of
5   terms are utilized. As used herein, the term "nucleic
acid" refers to a large molecule which can be single-
stranded or double-stranded, composed of monomers
(nucleotides) containing a sugar, phosphate and either a
purine or pyrimidine. A "nucleic acid fragment" is a
10 fraction of a given nucleic acid molecule. In higher
plants, deoxyribonucleic acid (DNA) is the genetic
material while ribonucleic acid (RNA) is involved in the
transfer of the information in DNA into proteins. A
"genome" is the entire body of genetic material
15 contained in each cell of an organism. The term
"nucleotide sequence" refers to a polymer of DNA or RNA
which can be single- or double-stranded, optionally
containing synthetic, non-natural or altered nucleotide
bases capable of incorporation into DNA or RNA polymers.
20   "Gene" refers to a nucleic acid fragment that
expresses a specific protein, including regulatory
sequences preceding (5' non-coding) and following (3'
non-coding) the coding region. "Native" gene refers to
the gene as found in nature with its own regulatory
25 sequences. "Chimeric" gene refers to a gene comprising
heterogeneous regulatory and coding sequences.
"Endogenous" gene refers to the native gene normally
found in its natural location in the genome. A
"foreign" gene refers to a gene not normally found in
30 the host organism but that is introduced by gene
transfer.

"Coding sequence" refers to a DNA sequence that
codes for a specific protein and excludes the non-coding
sequences.

14

"Initiation codon" and "termination codon" refer to
a unit of three adjacent nucleotides in a coding
sequence that specifies initiation and chain
termination, respectively, of protein synthesis (mRNA

5   translation). "Open reading frame" refers to the amino
acid sequence encoded between translation initiation and
termination codons of a coding sequence.
As used herein, suitable "regulatory sequences"
refer to nucleotide sequences located upstream (5'),

10   within, and/or downstream (3') to a coding sequence,
which control the transcription and/or expression of the
coding sequences, potentially in conjunction with the
protein biosynthetic apparatus of the cell. These
regulatory sequences include promoters, translation

15   leader sequences, transcription termination sequences,
and polyadenylation sequences.
"Promoter" refers to a DNA sequence in a gene,
usually upstream (5') to its coding sequence, which
controls the expression of the coding sequence by

20   providing the recognition for RNA polymerase and other
factors required for proper transcription. A promoter
may also contain DNA sequences that are involved in the
binding of protein factors which control the
effectiveness of transcription initiation in response to

25   physiological or developmental conditions. It may also
contain enhancer elements.
An "enhancer" is a DNA sequence which can stimulate
promoter activity. It may be an innate element of the
promoter or a heterologous element inserted to enhance

30   the level and/or tissue-specificity of a promoter.
"Constitutive promoters" refers to those that direct
gene expression in all tissues and at all times.
"Organ-specific" or "development-specific" promoters as
referred to herein are those that direct gene expression

35   almost exclusively in specific organs, such as leaves or

15

seeds, or at specific development stages in an organ,
such as in early or late embryogenesis, respectively.

The term "operably linked" refers to nucleic acid
sequences on a single nucleic acid molecule which are
5   associated so that the function of one is affected by
the other. For example, a promoter is operably linked
with a structure gene when it is capable of affecting
the expression of that structural gene (i.e., that the
structural gene is under the transcriptional control of
10   the promoter).

The term "expression", as used herein, is intended
to mean the production of the protein product encoded by
a gene. More particularly, "expression" refers to the
transcription and stable accumulation of the sense
15   (mRNA) or tha antisense RNA derived from the nucleic
acid fragment(s) of the invention that, in conjuction
with the protein apparatus of the cell, results in
altered levels of protein product. "Antisense
inhibition" refers to the production of antisense RNA
20   transcripts capable of preventing the expression of the
target protein. "Overexpression" refers to the
production of a gene product in transgenic organisms
that exceeds levels of production in normal or non-
transformed organisms. "Cosuppression" refers to the
25   expression of a foreign gene which has substantial
homology to an endogenous gene resulting in the
suppression of expression of both the foreign and the
endogenous gene. "Altered levels" refers to the
production of gene product(s) in transgenic organisms in
30   amounts or proportions that differ from that of normal
or non-transformed organisms.

The "3' non-coding sequences" refers to the DNA
sequence portion of a gene that contains a
polyadenylation signal and any other regulatory signal
35   capable of affecting mRNA processing or gene expression.

16

The polyadenylation signal is usually characterized by
affecting the addition of polyadenylic acid tracts to
the 3' end of the mRNA precursor.

The "translation leader sequence" refers to that
5   DNA sequence portion of a gene between the promoter and
coding sequence that is transcribed into RNA and is
present in the fully processed mRNA upstream (5') of the
translation start codon. The translation leader
sequence may affect processing of the primary transcript
10   to mRNA, mRNA stability or translation efficiency.

"Mature" protein refers to a post-translationally
processed polypeptide without its targeting signal.
"Precursor" protein refers to the primary product of
translation of mRNA. A "chloroplast targeting signal"
15   is an amino acid sequence which is translated in
conjunction with a protein and directs it to the
chloroplast. "Chloroplast transit sequence" refers to a
nucleotide sequence that encodes a chloroplast targeting
signal.

20   "Transformation" herein refers to the transfer of a
foreign gene into the genome of a host organism and its
genetically stable inheritance. Examples of methods of
plant transformation include Agrobacterium-mediated
transformation and particle-accelerated or "gene gun"
25   transformation technology.

"Amino acids" herein refer to the naturally
occuring L amino acids (Alanine, Arginine, Aspartic
acid, Asparagine, Cystine, Glutamic acid, Glutamine,
Glycine, Histidine, Isoleucine, Leucine, Lysine,
30   Methionine, Proline, Phenylalanine, Serine, Threonine,
Tryptophan, Tyrosine, and Valine). "Essential amino
acids" are those amino acids which cannot be synthesized
by animals. A "polypeptide" or "protein" as used herein
refers to a molecule composed of monomers (amino acids)

17

linearly linked by amide bonds (also known as peptide
bonds).

"Synthetic protein" herein refers to a protein
consisting of amino acid sequences that are not known to
5   occur in nature. The amino acid sequence may be derived
from a consensus of naturally occuring proteins or may
be entirely novel.

"Primary sequence" refers to the connectivity order
of amino acids in a polypeptide chain without regard to
10   the conformation of the molecule. Primary sequences are
written from the amino terminus to the carboxy terminus
of the polypeptide chain by convention.

"Secondary structure" herein refers to physico-
chemically favored regular backbone arrangements of a
15   polypeptide chain without regard to variations in side
chain identities or conformations. "Alpha helices" as
used herein refer to right-handed helices with
approximately 3.6 residues residues per turn of the
helix. An "amphipathic helix" refers herein to a
20   polypeptide in a helical conformation where one side of
the helix is predominantly hydrophobic and the other
side is predominantly hydrophilic.

"Coiled-coil" herein refers to an aggregate of two
parallel right-handed alpha helices which are wound
25   around each other to form a left-handed superhelix.

"Salt bridges" as discussed here refer to acid-base
pairs of charged amino acid side chains so arranged in
space that an attractive electrostatic interaction is
maintained between two parts of a polypeptide chain or
30   between one chain and another.

"Host cell" means the cell that is transformed with
the introduced genetic material.

                    Isolation of DHDPS genes
The E. coli dapA gene (ecodapA) was obtained as a
35   bacteriophage lambda clone from an ordered library of

18

3400 overlapping segments of <u>E. coli</u> DNA constructed by
Kohara, Akiyame and Isono [Kohara et al. (1987) Cell
50:595-508]. Details of the isolation and modification
of eco<u>dapA</u> are presented in Example 1. The eco<u>dapA</u> gene
5   encodes a DHDPS enzyme that is at least 20-fold less
sensitive to inhibition by lysine than a typical plant
enzyme, e.g., wheat DHDPS. For purposes of the present
invention, 20-fold less sensitive to inhibition by
lysine is termed lysine-insensitive.
10   The <u>Corynebacterium dapA</u> gene (cor<u>dapA</u>) was
isolated from genomic DNA from ATCC strain 13032 using
polymerase chain reaction (PCR). The nucleotide
sequence of the <u>Corynebacterium dapA</u> gene has been
published [Bonnassie et al. (1990) Nucleic Acids Res.
15   18:6421]. From the sequence it was possible to design
oligonucleotide primers for polymerase chain reaction
(PCR) that would allow amplification of a DNA fragment
containing the gene, and at the same time add unique
restriction endonuclease sites at the start codon and
20   just past the stop codon of the gene to facilitate
further constructions involving the gene. The details
of the isolation of the <u>Corynebacterium dapA</u> (cor<u>dapA</u>)
gene are presented in Example 1. The cor<u>dapA</u> gene
encodes a preferred lysine-insensitive DHDPS enzyme that
25   is unaffected by the presence of 70mM lysine in the
enzyme reaction mix.
        The isolation of other genes encoding DHDPS has
been described in the literature. A cDNA encoding DHDPS
from wheat [Kaneko et al. (1990) J. Biol. Chem.
30   265:17451-17455], and a cDNA encoding DHDPS from corn
[Frisch et al. (1991) Mol. Gen. Genet. 228:287-293] are
two examples of plant DHDPS genes that have been
isolated and sequenced. The plant genes encode wild
type lysine-sensitive DHDPS enzymes. However, Negrutui
35   et al. [(1984) Theor. Appl. Genet. 68:11-20], obtained

two AEC-resistant tobacco mutants in which DHDPS
activity was less sensitive to lysine inhibition than
the wild type enzyme. This indicates that these tobacco
mutants contain DHDPS genes encoding lysine-resistant
5   enzyme. These genes could be readily isolated from the
tobacco mutants using the methods already described for
isolating the wheat or corn genes or, alternatively, by
using the wheat or corn genes as heterologous
hybridization probes.
10   Still other genes encoding DHDPS can be isolated by
using either the E. coli dapA gene, the cordapA gene, or
either of the plant DHDPS genes as DNA hybridization
probes. Alternatively, other genes encoding DHDPS could
be isolated by functional complementation of an E. coli
15   dapA mutant, as was done to isolate the cordapA gene
[Yeh et al. (1988) Mol. Gen. Genet. 212:105-111] and the
corn DHDPS gene.
          Construction of Chimeric Genes for Expression of
                   dapA Coding Region in Plants
20   The expression of foreign genes in plants is well-
established [De Blaere et al. (1987) Meth. Enzymol.
143:277-291]. Proper level of expression of dapA mRNA
may require the use of different chimeric genes
utilizing different promoters. Such chimeric genes can
25   be transferred into host plants either together in a
single expression vector or sequentially using more than
one vector. A preferred class of heterologous hosts for
the expression of the coding sequence of the dapA genes
are eukaryotic hosts, particularly the cells of higher
30   plants. Particularly preferred among the higher plants
and the seeds derived from them are rapeseed (Brassica
napus, B. campestris) and soybean (Glycine max).
          The origin of promoter chosen to drive the
expression of the coding sequence is not critical as
35   long as it has sufficient transcriptional activity to

20

accomplish the invention by expressing translatable mRNA
for dapA genes in the desired host tissue. Preferred
promoters are those that allow expression of the protein
specifically in seeds. This may be especially useful,
5   since seeds are the primary source of vegetable amino
acids and also since seed-specific expression will avoid
any potential deleterious effect in non-seed organs.
Examples of seed-specific promoters include, but are not
limited to, the promoters of seed storage proteins. The
10   seed storage proteins are strictly regulated, being
expressed almost exclusively in seeds in a highly organ-
specific and stage-specific manner [Higgins et al.(1984)
Ann. Rev. Plant Physiol. 35:191-221; Goldberg et
al.(1989) Cell 56:149-160; Thompson et al. (1989)
15   BioEssays 10:108-113]. Moreover, different seed storage
proteins may be expressed at different stages of seed
development.
    There are currently numerous examples for seed-
specific expression of seed storage protein genes in
20   transgenic dicotyledonous plants. These include genes
from dicotyledonous plants for bean β-phaseolin
[Sengupta-Goplalan et al. (1985) Proc. Natl. Acad. Sci.
USA 82:3320-3324; Hoffman et al. (1988) Plant Mol. Biol.
11:717-729], bean lectin [Voelker et al. (1987) EMBO J.
25   6: 3571-3577], soybean lectin [Okamuro et al. (1986)
Proc. Natl. Acad. Sci. USA 83:8240-8244], soybean kunitz
trypsin inhibitor [Perez-Grau et al. (1989) Plant Cell
1:095-1109], soybean β-conglycinin [Beachy et al. (1985)
EMBO J. 4:3047-3053; Barker et al. (1988) Proc. Natl.
30   Acad. Sci. USA 85:458-462; Chen et al. (1988) EMBO J.
7:297-302; Chen et al. (1989) Dev. Genet. 10:112-122;
Naito et al. (1988) Plant Mol. Biol. 11:109-123], pea
vicilin [Higgins et al. (1988) Plant Mol. Biol.
11:683-695], pea convicilin [Newbigin et al. (1990)
35   Planta 180:461], pea legumin [Shirsat et al. (1989) Mol.

21

Gen. Genetics 215:326]; rapeseed napin [Radke et al.
(1988) Theor. Appl. Genet. 75:685-694] as well as genes
from monocotyledonous plants such as for maize 15 kD
zein [Hoffman et al. (1987) EMBO J. 6:3213-3221;
5   Schernthaner et al. (1988) EMBO J. 7:1249-1253;
Williamson et al. (1988) Plant Physiol. 88:1002-1007],
barley β-hordein [Marris et al. (1988) Plant Mol. Biol.
10:359-366] and wheat glutenin [Colot et al. (1987) EMBO
J. 6:3559-3564]. Moreover, promoters of seed-specific
10 genes, operably linked to heterologous coding sequences
in chimeric gene constructs, also maintain their
temporal and spatial expression pattern in transgenic
plants. Such examples include Arabidopsis thaliana 2S
seed storage protein gene promoter to express enkephalin
15 peptides in Arabidopsis and B. napus seeds
[Vandekerckhove et al. (1989) Bio/Technology 7:929-932],
bean lectin and bean β-phaseolin promoters to express
luciferase [Riggs et al. (1989) Plant Sci. 63:47-57],
and wheat glutenin promoters to express chloramphenicol
20 acetyl transferase [Colot et al. (1987) EMBO J.
6:3559-3564].

Of particular use in the expression of the nucleic
acid fragment of the invention will be the promoters
from several extensively-characterized seed storage
25 protein genes such as those for bean β-phaseolin
[Sengupta-Goplalan et al. (1985) Proc. Natl. Acad. Sci.
USA 82:3320-3324; Hoffman et al. (1988) Plant Mol. Biol.
11:717-729], soybean Kunitz trypsin inhibitor [Jofuku et
al. (1989) Plant Cell 1:1079-1093; Perez-Grau et al.
30 (1989) Plant Cell 1:1095-1109], soybean β-conglycinin
[Harada et al. (1989) Plant Cell 1:415-425], and
rapeseed napin [Radke et al. (1988) Theor. Appl. Genet.
75:685-694]. Promoters of genes for bean β-phaseolin
and soybean β-conglycinin storage protein will be

22

particularly useful in expressing the dapA mRNA in the
cotyledons at mid- to late-stages of seed development.

Also of particular use in the expression of the
nucleic acid fragments of the invention will be the
5   heterologous promoters from several extensively
characterized corn seed storage protein genes such as
endosperm-specific promoters from the 10 kD zein
[Kirihara et al. (1988) Gene 71:359-370], the 27 kD zein
[Prat et al. (1987) Gene 52:51-49; Gallardo et al.
10   (1988) Plant Sci. 54:211-281; Reina et al. (1990)
Nucleic Acids Res. 18:6426-6426], and the 19 kD zein
[Marks et al. (1985) J. Biol. Chem. 260:16451-16459].
The relative transcriptional activities of these
promoters in corn have been reported [Kodrzyck et al.
15   (1989) Plant Cell 1:105-114] providing a basis for
choosing a promoter for use in chimeric gene constructs
for corn. For expression in corn embryos, the strong
embryo-specific promoter from the globulin 1 (GLB1) gene
[Kriz (1989) Biochemical Genetics 27:239-251, Wallace et
20   al. (1991) Plant Physiol. 95:973-975] can be used.

It is envisioned that the introduction of enhancers
or enhancer-like elements into other promoter constructs
will also provide increased levels of primary
transcription for dapA genes to accomplish the
25   invention. These would include viral enhancers such as
that found in the 35S promoter [Odell et al. (1988)
Plant Mol. Biol. 10:263-272], enhancers from the opine
genes [Fromm et al. (1989) Plant Cell 1:977-984], or
enhancers from any other source that result in increased
30   transcription when placed into a promoter operably
linked to the nucleic acid fragment of the invention.

Of particular importance is the DNA sequence
element isolated from the gene for the α'-subunit of
β-conglycinin that can confer 40-fold seed-specific
35   enhancement to a constitutive promoter [Chen et al.

23

(1988) EMBO J. 7:297-302; Chen et al. (1989) Dev. Genet.
10:112-122]. One skilled in the art can readily isolate
this element and insert it within the promoter region of
any gene in order to obtain seed-specific enhanced
5   expression with the promoter in transgenic plants.
Insertion of such an element in any seed-specific gene
that is expressed at different times than the
β-conglycinin gene will result in expression in
transgenic plants for a longer period during seed
10   development.

        Any 3' non-coding region capable of providing a
polyadenylation signal and other regulatory sequences
that may be required for the proper expression of the
dapA coding regions can be used to accomplish the
15   invention. This would include the 3' end from any
storage protein such as the 3' end of the bean phaseolin
gene, the 3' end of the soybean β-conglycinin gene, the
3' end from viral genes such as the 3' end of the 35S or
the 19S cauliflower mosaic virus transcripts, the 3' end
20   from the opine synthesis genes, the 3' ends of ribulose
1,5-bisphosphate carboxylase or chlorophyll a/b binding
protein, or 3' end sequences from any source such that
the sequence employed provides the necessary regulatory
information within its nucleic acid sequence to result
25   in the proper expression of the promoter/coding region
combination to which it is operably linked. There are
numerous examples in the art that teach the usefulness
of different 3' non-coding regions [for example, see
Ingelbrecht et al. (1989) Plant Cell 1:671-680].
30   DNA sequences coding for intracellular localization
sequences may be added to the dapA coding sequence if
required for the proper expression of the proteins to
accomplish the invention. Plant amino acid biosynthetic
enzymes are known to be localized in the chloroplasts
35   and therefore are synthesized with a chloroplast

24

targeting signal. Bacterial proteins such as
Corynebacterium DHDPS have no such signal. A
chloroplast transit sequence could, therefore, be fused
to the dapA coding sequence. Preferred chloroplast

5   transit sequences are those of the small subunit of
ribulose 1,5-bisphosphate carboxylase, e.g. from soybean
[Berry-Lowe et al. (1982) J. Mol. Appl. Genet.
1:483-498] for use in dicotyledonous plants and from
corn [Lebrun et al. (1987) Nucleic Acids Res. 15:4360]

10   for use in monocotyledonous plants.

<u>Introduction of dapA</u>
<u>Chimeric Genes into Plants</u>

       Various methods of introducing a DNA sequence
(i.e., of transforming) into eukaryotic cells of higher

15   plants are available (see EPO publications 0 295 959 A2
and 0 138 341 A1). Such methods include those based on
transformation vectors based on the Ti and Ri plasmids
of Agrobacterium spp. It is particularly preferred to
use the binary type of these vectors. Ti-derived

20   vectors transform a wide variety of higher plants,
including monocotyledonous and dicotyledonous plants,
such as soybean, cotton and rape [Pacciotti et al.
(1985) Bio/Technology 3:241; Byrne et al. (1987) Plant
Cell, Tissue and Organ Culture 8:3; Sukhapinda et al.

25   (1987) Plant Mol. Biol. 8:209-216; Lorz et al. (1985)
Mol. Gen. Genet. 199:178; Potrykus (1985) Mol. Gen.
Genet. 199:183].

       For introduction into plants the chimeric genes of
the invention can be inserted into binary vectors as

30   described in Examples 6-12. The vectors are part of a
binary Ti plasmid vector system [Bevan, (1984) Nucl.
Acids. Res. 12:8711-8720] of Agrobacterium tumefaciens.

       Other transformation methods are available to those
skilled in the art, such as direct uptake of foreign DNA

35   constructs [see EPO publication 0 295 959 A2],

techniques of electroporation [see Fromm et al. (1986)
Nature (London) 319:791] or high-velocity ballistic
bombardment with metal particles coated with the nucleic
acid constructs [see Kline et al. (1987) Nature (London)
5   327:70, and see U.S. Pat. No. 4,945,050]. Once
transformed, the cells can be regenerated by those
skilled in the art.

Of particular relevance are the recently described
methods to transform foreign genes into commercially
10   important crops, such as rapeseed [see De Block et al.
(1989) Plant Physiol. 91:694-701], sunflower [Everett
et al. (1987) Bio/Technology 5:1201], soybean [McCabe
et al. (1988) Bio/Technology 6:923; Hinchee et al.
(1988) Bio/Technology 6:915; Chee et al. (1989) Plant
15   Physiol. 91:1212-1218; Christou et al. (1989) Proc.
Natl. Acad. Sci USA 86:7500-7504; EPO Publication
0 301 749 A2], and corn [Gordon-Kamm et al. (1990) Plant
Cell 2:603-618; Fromm et al. (1990) Biotechnology
8:833-839].

20   For introduction into plants by high-velocity
ballistic bombardment, the chimeric genes of the
invention can be inserted into suitable vectors as
described in Example 6.

                  Expression of dapA Chimeric Genes in
25   Rapeseed, Soybean and Corn Plants

To analyze for expression of the chimeric dapA gene
in seeds and for the consequences of expression on the
amino acid content in the seeds, a seed meal can be
prepared as described in Examples 5 or 6 or by any other
30   suitable method. The seed meal can be partially or
completely defatted, via hexane extraction for example,
if desired. Protein extracts can be prepared from the
meal and analyzed for DHDPS enzyme activity.
Alternatively the presence of the DHDPS protein can be
35   tested for immunologically by methods well-known to

26

those skilled in the art. Nearly all of the
transformants expressed the foreign DHDPS protein (see
Examples 5, 6 and 13). To measure free amino acid
composition of the seeds, free amino acids can be
5   extracted from the meal and analyzed by methods known to
those skilled in the art (see Examples 5 and 6 for
suitable procedures).

Rapeseed transformants expressing DHDPS protein
showed a greater than 100-fold increase in free lysine
10 level in their seeds. There was a good correlation
between transformants expressing higher levels of DHDPS
protein and those having higher levels of free lysine.
Among the transformants, there has been no greater
accumulation of free lysine due to expression of a
15 lysine insensitive AK enzyme along with a lysine-
insensitive DHDPS compared to expression of a lysine-
insensitive DHDPS alone. Thus, in rapeseed, expression
of a lysine-insensitive DHDPS in seeds is necessary and
sufficient to cause a large increase in free lysine. A
20 high level of α-aminoadipic acid, indicative of lysine
catabolism, was observed in all of the transformed lines
with increased levels of free lysine.

To measure the total amino acid composition of
mature rapeseed seeds, defatted meal was analyzed as
25 described in Example 5. Relative amino acid levels in
the seeds were compared as percentages of lysine to
total amino acids. The highest expressing lines showed
a nearly 2-fold increase in the lysine level in the
seeds, so that lysine makes up about 12% of the total
30 seed amino acids.

Twenty-one of twenty-three soybean transformants
expressed the DHDPS protein. Analysis of single seeds
of these transformants showed excellent correlation
between expression of the GUS transformation marker gene
35 and DHDPS in individual seeds. Therefore, the GUS and

27

DHDPS genes are integrated at the same site in the soybean genome.

There was excellent correlation between transformants expressing _Corynebacteria_ DHDPS protein
5   and those having higher levels of free lysine. From 20-fold to 120-fold increases in free lysine level was observed in seeds expressing _Corynebacteria_ DHDPS.

Analyses of free lysine levels in individual seeds from transformants in which the transgenes segregated as
10 a single locus revealed that the increase in free lysine level was significantly higher in about one-fourth of the seeds. Since one-fourth of the seeds are expected to be homozygous for the transgene, it is likely that the higher lysine seeds are the homozygotes. Further-
15 more, this indicates that the level of increase in free lysine is dependent upon the copy number of the DHDPS gene. Therefore, lysine levels could be further increased by making hybrids of two different transformants, and obtaining progeny that are homozygous
20 at both transgene loci, thus increasing the copy number of the DHDPS gene from two to four.

A high level of saccharopine, indicative of lysine catabolism, was observed in seeds that contained high levels of lysine. Thus, prevention of lysine catabolism
25 by inactivation of lysine ketoglutarate reductase should further increase the accumulation of free lysine in the seeds. Alternatively, incorporation of lysine into a peptide or lysine-rich protein would prevent catabolism and lead to an increase in the accumulation of lysine in
30 the seeds.

Total lysine levels were significantly increased in seeds expressing _Corynebacteria_ DHDPS protein. Seeds with a 10-260% increase in the lysine level compared to the untransformed control were observed . Expression of
35 DHDPS along with a lysine-insensitive aspartokinase

28

enzyme resulted in lysine increases of more than 400%.
Thus, these seeds contain much more lysine than any
previous soybean seed.

Expression of the <u>Corynebacterium</u> DHDPS protein,
5   driven by either the corn globulin 1 promoter for
expression in the embryo or the corn glutelin 2 promoter
for expression in the endosperm, was observed in the
corn seeds. Free lysine levels in the seeds increased
from about 1.4% of free amino acids in control seeds to
10   15-27% of free amino acids in seeds expressing
<u>Corynebacterium</u> DHDPS from the globulin 1 promoter. A
smaller increase in free lysine was observed in in seeds
expressing <u>Corynebacterium</u> DHDPS from the glutelin 2
promoter. Thus to increase lysine, it may be better to
15   express this enzyme in the embryo rather than the
endosperm. A high level of saccharopine, indicative of
lysine catabolism, was observed in seeds that contained
high levels of lysine. The increased accumulation of
free lysine in seeds expressing <u>Corynebacterium</u> DHDPS
20   from the globulin 1 promoter was sufficient to result in
substantial increases (35%-130%) in the total lysine
content of the seeds.

<div align="center"><u>Isolation of a Plant</u>

<u>Lysine Ketoglutarate Reductase Gene</u></div>

25   To accumulate higher levels of free lysine it may
be desirable to prevent lysine catabolism. Evidence
indicates that lysine is catabolized in plants via the
saccharopine pathway. The first enzymatic evidence for
the existence of this pathway was the detection of
30   lysine ketoglutarate reductase (LKR) activity in
immature endosperm of developing maize seeds [Arruda et
al. (1982) Plant Physiol. 69:988-989]. LKR catalyzes
the first step in lysine catabolism, the condensation of
L-lysine with α-ketoglutarate into saccharopine using
35   NADPH as a cofactor. LKR activity increases sharply

29

from the onset of endosperm development in corn, reaches
a peak level at about 20 days after pollination, and
then declines [Arruda et al. (1983) Phytochemistry
22:2687-2689]. In order to prevent the catabolism of
5   lysine it would be desirable to reduce or eliminate LKR
expression or activity. This could be accomplished by
cloning the LKR gene, preparing a chimeric gene for
cosuppression of LKR or preparing a chimeric gene to
express antisense RNA for LKR, and introducing the
10   chimeric gene into plants via transformation.

Several methods to clone a plant LKR gene are
available to one skilled in the art. The protein can be
purified from corn endosperm, as described in Brochetto-
Braga et al. [(1992) Plant Physiol. 98:1139-1147] and
15   used to raise antibodies. The antibodies can then be
used to screen an cDNA expression library for LKR
clones. Alternatively the purified protein can be used
to determine amino acid sequence at the amino-terminal
of the protein or from protease derived internal peptide
20   fragments. Degenerate oligonucleotide probes can be
prepared based upon the amino acid sequence and used to
screen a plant cDNA or genomic DNA library via
hybridization. Another method makes use of an E. coli
strain that is unable to grow in a synthetic medium
25   containing 20 μg/mL of L-lysine. Expression of LKR
full-length cDNA in this strain will reverse the growth
inhibition by reducing the lysine concentration.
Construction of a suitable E. coli strain and its use to
select clones from a plant cDNA library that lead to
30   lysine-resistant growth is described in Example 7.

In order to block expression of the LKR gene in
transformed plants, a chimeric gene designed for
cosuppression of LKR can be constructed by linking the
LKR gene or gene fragment to any of the plant promoter
35   sequences described above (U.S. Patent No. 5,231,020).

Alternatively, a chimeric gene designed to express
antisense RNA for all or part of the LKR gene can be
constructed by linking the LKR gene or gene fragment in
reverse orientation to any of the plant promoter

5   sequences described above (Eur. Patent Applic.
No. 84112647.7). Either the cosuppression or antisense
chimeric gene could be introduced into plants via
transformation. Transformants wherein expression of the
endogenous LKR gene is reduced or eliminated are

10   selected.

Preferred promoters for the chimeric genes would be
seed-specific promoters. For soybean, rapeseed and
other dicotyledonous plants, strong seed-specific
promoters from a bean phaseolin gene, a soybean

15   β-conglycinin gene, glycinin gene, Kunitz trypsin
inhibitor gene, or rapeseed napin gene would be
preferred. For corn and other monocotyledonous plants,
a strong endosperm-specific promoter, e.g., the 10 kD or
27 kD zein promoter, would be preferred.

20   Transformed plants containing any of the chimeric
LKR genes can be obtained by the methods described
above. In order to obtain transformed plants that
express a chimeric gene for cosuppression of LKR or
antisense LKR, as well as a chimeric gene encoding

25   lysine-insensitive DHDPS, the cosuppression or antisense
LKR gene could be linked to the chimeric gene encoding
lysine-insensitive DHDPS and the two genes could be
introduced into plants via transformation.
Alternatively, the chimeric gene for cosuppression of

30   LKR or antisense LKR could be introduced into previously
transformed plants that express lysine-insensitive
DHDPS, or the cosuppression or antisense LKR gene could
be introduced into normal plants and the transformants
obtained could be crossed with plants that express

35   lysine-insensitive DHDPS.

Design of Lysine-Rich Polypeptides

It may be desirable to convert the high levels of
lysine produced into a form that is insensitive to
breakdown, e.g., by incorporating it into a di-, tri- or
5   oligopeptide, or a lysine-rich storage protein. No
natural lysine-rich proteins are known.

One aspect of this invention is the design of
polypeptides which can be expressed in vivo to serve as
lysine-rich seed storage proteins. Polypeptides are
10   linear polymers of amino acids where the α-carboxyl
group of one amino acid is covalently bound to the
α-amino group of the next amino acid in the chain. Non-
covalent interactions among the residues in the chain
and with the surrounding solvent determine the final
15   conformation of the molecule. Those skilled in the art
must consider electrostatic forces, hydrogen bonds,
Van der Waals forces, hydrophobic interactions, and
conformational preferences of individual amino acid
residues in the design of a stable folded polypeptide
20   chain [see for example: Creighton, (1984) Proteins,
Structures and Molecular Properties, W. H. Freeman and
Company, New York, pp. 133-197, or Schulz et al., (1979)
Principles of Protein Structure, Springer Verlag, New
York, pp. 27-45]. The number of interactions and their
25   complexity suggest that the design process may be aided
by the use of natural protein models where possible.

The synthetic storage proteins (SSPs) embodied in
this invention are chosen to be polypeptides with the
potential to be enriched in lysine relative to average
30   levels of proteins in plant seeds. Lysine is a charged
amino acid at physiological pH and is therefore found
most often on the surface of protein molecules [Chotia,
(1976) Journal of Molecular Biology 105:1-14]. To
maximize lysine content, Applicants chose a molecular
35   shape with a high surface-to-volume ratio for the

synthetic storage proteins embodied in this invention.
The alternatives were either to stretch the common
globular shape of most proteins to form a rod-like
extended structure or to flatten the globular shape to a
5   disk-like structure. Applicants chose the former
configuration as there are several natural models for
long rod-like proteins in the class of fibrous proteins
[Creighton, (1984) Proteins, Structures and Molecular
Properties, W.H. Freeman and Company, New York, p. 191].
10   Coiled-coils constitute a well-studied subset of
the class of fibrous proteins [see Cohen et al., (1986)
Trends Biochem. Sci. 11:245-248]. Natural examples are
found in α-keratins, paramyosin, light meromyosin and
tropomyosin. These protein molecules consist of two
15   parallel alpha helices twisted about each other in a
left-handed supercoil. The repeat distance of this
supercoil is 140 Å (compared to a repeat distance of
5.4 Å for one turn of the individual helices). The
supercoil causes a slight skew (10°) between the axes of
20   the two individual alpha helices.
         In a coiled coil there are 3.5 residues per turn of
the individual helices resulting in an exact 7 residue
periodicity with respect to the superhelix axis (see
Figure 1). Every seventh amino acid in the polypeptide
25   chain therefore occupies an equivalent position with
respect to the helix axis. Applicants refer to the
seven positions in this heptad unit of the invention as
(d e f g a b c) as shown in Figures 1 and 2a. This
conforms to the conventions used in the coiled-coil
30   literature.
         The a and d amino acids of the heptad follow a 4,3
repeat pattern in the primary sequence and fall on one
side of an individual alpha helix (See Figure 1). If
the amino acids on one side of an alpha helix are all
35   non-polar, that face of the helix is hydrophobic and

33

will associate with other hydrophobic surfaces as, for
example, the non-polar face of another similar helix. A
coiled-coil structure results when two helices dimerize
such that their hydrophobic faces are aligned with each
5   other (See Figure 2a).

The amino acids on the external faces of the
component alpha helices (b, c, e, f, g) are usually
polar in natural coiled-coils in accordance with the
expected pattern of exposed and buried residue types in
10   globular proteins [Schulz, et al., (1979) Principles of
Protein Structure. Springer Verlag, New York, p. 12;
Talbot, et al , (1982) Acc. Chem. Res. 15:224-230;
Hodges et al., (1981) Journal of Biological Chemistry
256:1214-1224]. Charged amino acids are sometimes found
15   forming salt bridges between positions e and g' or
positions g and e' on the opposing chain (see
Figure 2a).

Thus, two amphipathic helices like the one shown in
Figure 1 are held together by a combination of
20   hydrophobic interactions between the a, a', d, and d'
residues and by salt bridges between e and g' and/or g
and e' residues. The packing of the hydrophobic
residues in the supercoil maintains the chains "in
register". For short polypeptides comprising only a few
25   turns of the component alpha helical chains, the 10°
skew between the helix axes can be ignored and the two
chains treated as parallel (as shown in Figure 2a).

A number of synthetic coiled-coils have been
reported in the literature (Lau et al., (1984) Journal
30   of Biological Chemistry 259:13253-13261; Hodges et al.,
(1988) Peptide Research 1:19-30; DeGrado et al., (1989)
Science 243:622-628; O'Neil et al., (1990) Science
250:646-651]. Although these polypeptides vary in size,
Lau et al. found that 29 amino acids were sufficient for
35   dimerization to form the coiled-coil structure [Lau et

34

al., (1984) Journal of Biological Chemistry
259:13253-13261]. Applicants constructed the
polypeptides in this invention as 28-residue and larger
chains for reasons of conformational stability.

5   The polypeptides of this invention are designed to
dimerize with a coiled-coil motif in aqueous
environments. Applicants have used a combination of
hydrophobic interactions and electrostatic interactions
to stabilize the coiled-coil conformation. Most
10   nonpolar residues are restricted to the a and d
positions which creates a hydrophobic stripe parallel to
the axis of the helix. This is the dimerization face.
Applicants avoided large, bulky amino acids along this
face to minimize steric interference with dimerization
15   and to facilitate formation of the stable coiled-coil
structure.

        Despite recent reports in the literature suggesting
that methionine at positions a and d is destabilizing to
coiled-coils in the leucine zipper subgroup [Landschulz
20   et al., (1989) Science 243:1681-1688 and Hu et al.,
(1990) Science 250:1400-1403], Applicants chose to
substitute methionine residues for leucine on the
hydrophobic face of the SSP polypeptides. Methionine
and leucine are similar in molecular shape (Figure 3).
25   Applicants demonstrated that any destabilization of the
coiled-coil that may be caused by methionine in the
hydrophobic core appears to be compensated in sequences
where the formation of salt bridges (e-g' and g-e')
occurs at all possible positions in the helix (i.e.,
30   twice per heptad).

        To the extent that it is compatible with the goal
of creating a polypeptide enriched in lysine, Applicants
minimized the unbalanced charges in the polypeptide.
This may help to prevent undesirable interactions

35

between the synthetic storage proteins and other plant
proteins when the polypeptides are expressed _in vivo_.

    The polypeptides of this invention are designed to
spontaneously fold into a defined, conformationally

5   stable structure, the alpha helical coiled-coil, with
minimal restrictions on the primary sequence. This
allows synthetic storage proteins to be custom-tailored
for specific end-user requirements. Any amino acid can
be incorporated at a frequency of up to one in every

10   seven residues using the b, c, and f positions in the
heptad repeat unit. Applicants note that up to 43% of
an essential amino acid from the group isoleucine,
leucine, lysine, methionine, threonine, and valine can
be incorporated and that up to 14% of the essential

15   amino acids from the group phenylalanine, tryptophan,
and tyrosine can be incorporated into the synthetic
storage proteins of this invention.

    In the SSPs only Met, Leu, Ile, Val or Thr are
located in the hydrophobic core. Furthermore, the e, g,

20   e', and g' positions in the SSPs are restricted such
that an attractive electrostatic interaction always
occurs at these positions between the two polypeptide
chains in an SSP dimer. This makes the SSP polypeptides
more stable as dimers.

25   Thus, the novel synthetic storage proteins
described in this invention represent a particular
subset of possible coiled-coil polypeptides. Not all
polypeptides which adopt an amphipathic alpha helical
conformation in aqueous solution are suitable for the

30   applications described here.

    The following rules derived from Applicants' work
define the SSP polypeptides that Applicants use in their
invention:

The synthetic polypeptide comprises n heptad units
(d e f g a b c), each heptad being either the same or
different, wherein:

    n is at least 4;

    a and d are independently selected from the
        group consisting of Met, Leu, Val, Ile and
        Thr;

    e and g are independently selected from the
        group consisting of the acid/base pairs
        Glu/Lys, Lys/Glu, Arg/Glu, Arg/Asp,
        Lys/Asp, Glu/Arg, Asp//Arg and Asp/Lys;
        and

    b, c and f are independently any amino acids
        except Gly or Pro and at least two amino
        acids of b, c and f in each heptad are
        selected from the group consisting of Glu,
        Lys, Asp, Arg, His, Thr, Ser, Asn, Gln,
        Cys and Ala.

Chimeric Genes Encoding Lysine-Rich Polypeptides

DNA sequences which encode the polypeptides
described above can be designed based upon the genetic
code. Where multiple codons exist for particular amino
acids, codons should be chosen from those preferable for
translation in plants. Oligonucleotides corresponding
to these DNA sequences can be synthesized using an ABI
DNA synthesizer, annealed with oligonucleotides
corresponding to the complementary strand and inserted
into a plasmid vector by methods known to those skilled
in the art. The encoded polypeptide sequences can be
lengthened by inserting additional annealed oligonucleo-
tides at restriction endonuclease sites engineered into
the synthetic gene. Some representative strategies for
constructing genes encoding lysine-rich polypeptides of
the invention, as well as DNA and amino acid sequences
of preferred embodiments are provided in Example 8.

A chimeric gene designed to express RNA for a
synthetic storage protein gene encoding a lysine-rich
polypeptide can be constructed by linking the gene to
any of the plant promoter sequences described above.

5 Preferred promoters would be seed-specific promoters.
For soybean, rapeseed and other dicotyledonous plants
strong seed-specific promoters from a bean phaseolin
gene, a soybean β-conglycinin gene, glycinin gene,
Kunitz trypsin inhibitor gene, or rapeseed napin gene

10 would be preferred. For corn or other monocotyledonous
plants, a strong endosperm-specific promoter, e.g., the
10 kD or 27 kD zein promoter, or a strong embyro-
specific promoter, e.g., the corn globulin 1 promoter,
would be preferred.

15   In order to obtain plants that express a chimeric
gene for a synthetic storage protein gene encoding a
lysine-rich polypeptide, plants can be transformed by
any of the methods described above. In order to obtain
plants that express both a chimeric SSP gene and a

20 chimeric gene encoding lysine-insensitive DHDPS, the SSP
gene could be linked to the chimeric gene encoding
lysine-insensitive DHDPS and the two genes could be
introduced into plants via transformation.
Alternatively, the chimeric SSP gene could be introduced

25 into previously transformed plants that express lysine-
insensitive DHDPS, or the SSP gene could be introduced
into normal plants and the transformants obtained could
be crossed with plants that express lysine-insensitive
DHDPS.

30   Results from genetic crosses of transformed plants
containing lysine biosynthesis genes with transformed
plants containing lysine-rich protein genes (see
Example 10) demonstrate that the total lysine levels in
seeds can be increased by the coordinate expression of

35 these genes. This result was especially striking

38

because the gene copy number of all of the transgenes
was reduced in the hybrid. It is expected that the
lysine level would be further increased if the
biosynthesis genes and the lysine-rich protein genes
5   were all homozygous.

<center>EXAMPLES</center>

The present invention is further defined in the
following Examples, in which all parts and percentages
are by weight and degrees are Celsius, unless otherwise
10   stated.

<center>EXAMPLE 1</center>

<center>Isolation of the E. coli and Corynebacterium
glutamicum dapA genes</center>

The E. coli dapA gene (ecodapA) has been cloned,
15   restriction endonuclease mapped and sequenced previously
[Richaud et al. (1986) J. Bacteriol. 166:297-300]. For
the present invention the dapA gene was obtained on a
bacteriophage lambda clone from an ordered library of
3400 overlapping segments of cloned E. coli DNA
20   constructed by Kohara, Akiyama and Isono [Kohara et al.
(1987) Cell 50:595-508]. From the knowledge of the map
position of dapA at 53 min on the E. coli genetic map
[Bachman (1983) Microbiol. Rev. 47:180-230], the
restriction endonuclease map of the cloned gene [Richaud
25   et al. (1986) J. Bacteriol. 166:297-300], and the
restriction endonuclease map of the cloned DNA fragments
in the E. coli library [Kohara et al. (1987) Cell
50:595-508], it was possible to choose lambda phages
4C11 and 5A8 [Kohara et al. (1987) Cell 50:595-508] as
30   likely candidates for carrying the dapA gene. The
phages were grown in liquid culture from single plaques
as described [see Current Protocols in Molecular Biology
(1987) Ausubel et al. eds., John Wiley & Sons New York]
using LE392 as host [see Sambrook et al. (1989)
35   Molecular Cloning: a Laboratory Manual, Cold Spring

Harbor Laboratory Press]. Phage DNA was prepared by
phenol extraction as described [see Current Protocols in
Molecular Biology (1987) Ausubel et al. eds., John Wiley
& Sons, New York]. Both phages contained an

5   approximately 2.8 kb Pst I DNA fragment expected for the
dapA gene [Richaud et al. (1986) J. Bacteriol.
166:297-300]. The fragment was isolated from the digest
of phage 5A8 and inserted into Pst I digested vector
pBR322 yielding plasmid pBT427.

10   The Corynebacterium dapA gene (cordapA) was
isolated from genomic DNA from ATCC strain 13032 using
polymerase chain reaction (PCR). The nucleotide
sequence of the Corynebacterium dapA gene has been
published [Bonnassie et al. (1990) Nucleic Acids Res.

15   18:6421]. From the sequence it was possible to design
oligonucleotide primers for PCR that would allow
amplification of a DNA fragment containing the gene, and
at the same time add unique restriction endonuclease
sites at the start codon (Nco I) and just past the stop

20   codon (EcoR I) of the gene. The oligonucleotide primers
used were:

     SEQ ID NO:1:
     CCCGGGCCAT GGCTACAGGT TTAACAGCTA AGACCGGAGT AGAGCACT
25

     SEQ ID NO:2:
     GATATCGAAT TCTCATTATA GAACTCCAGC TTTTTTC

        PCR was performed using a Perkin-Elmer Cetus kit
30   according to the instructions of the vendor on a
     thermocycler manufactured by the same company. The
     reaction product, when run on an agarose gel and stained
     with ethidium bromide, showed a strong DNA band of the
     size expected for the Corynebacterium dapA gene, about

35   900 bp. The PCR-generated fragment was digested with
     restriction endonucleases Nco I and EcoR I and inserted

40

into expression vector pBT430 (see Example 2) digested
with the same enzymes. In addition to introducing an
Nco I site at the translation start codon, the PCR
primers also resulted in a change of the second codon

5   from AGC coding for serine to GCT coding for alanine.
Several clones that expressed active, lysine-insensitive
DHDPS (see Example 2) were isolated, indicating that the
second codon amino acid substitution did not affect
activity; one clone was designated FS766.

10   The Nco I to EcoR I fragment carrying the
PCR-generated Corynebacterium dapA gene was subcloned
into the phagemid vector pGEM-9Zf(-) from Promega,
single-stranded DNA was prepared and sequenced. This
sequence is shown in SEQ ID NO:3.

15   Aside from the differences in the second codon
already mentioned, the sequence matched the published
sequence except at two positions, nucleotides 798 and
799. In the published sequence these are TC, while in
the gene shown in SEQ ID NO:3 they are CT. This change

20   results in an amino acid substitution of leucine for
serine. The reason for this difference is not known.
It may be due to an error in the published sequence, the
difference in strains used to isolate the gene, or a
PCR-generated error. The latter seems unlikely since

25   the same change was observed in at least 3 independently
isolated PCR-generated dapA genes. The difference has
no apparent effect on DHDPS enzyme activity (see
Example 2).

                          EXAMPLE 2

30   High level expression of the E. coli and
            Corynebacterium glutamicum dapA genes in E. coli
          An Nco I (CCATGG) site was inserted at the
translation initiation codon of the E. coli dapA gene
using oligonucleotide-directed mutagenesis. The 2.8 kb

35   Pst I DNA fragment carrying the dapA gene in plasmid

41

pBT427 (see Example 1) was inserted into the Pst I site
of phagemid vector pTZ18R (Pharmacia) yielding pBT431.
The orientation of the dapA gene was such that the
coding strand would be present on the single-stranded
5   phagemid DNA. Oligonucleotide-directed mutagenesis was
carried out using a Muta-Gene kit from Bio-Rad according
to the manufacturer's protocol with the mutagenic primer
shown below:

10   SEQ ID NO:4:
CTTCCCGTGA CCATGGGCCA TC

Putative mutants were screened for the presence of an
Nco I site and a plasmid, designated pBT437, was shown
15   to have the the proper sequence in the vicinity of the
mutation by DNA sequencing. The addition of an Nco I
site at the translation start codon also resulted in a
change of the second codon from TTC coding for
phenylalanine to GTC coding for valine.
20   To achieve high level expression of the dapA genes
in E. coli the bacterial expression vector pBT430. This
expression vector is a derivative of pET-3a [Rosenberg
et al. (1987) Gene 56:125-135] which employs the
bacteriophage T7 RNA polymerase/T7 promoter system.
25   Plasmid pBT430 was constructed by first destroying the
EcoR I and Hind III sites in pET-3a at their original
positions. An oligonucleotide adaptor containing EcoR I
and Hind III sites was inserted at the BamH I site of
pET-3a. This created pET-3aM with additional unique
30   cloning sites for insertion of genes into the expression
vector. Then, the Nde I site at the position of
translation initiation was converted to an Nco I site
using oligonucleotide-directed mutagenesis. The DNA
sequence of pET-3aM in this region, 5'-CATATGG, was
35   converted to 5'-CCCATGG in pBT430.

42

The E. coli dapA gene was cut out of plasmid pBT437
as an 1150 bp Nco I–Hind III fragment and inserted into
the expression vector pBT430 digested with the same
enzymes, yielding plasmid pBT442. For expression of the
5   Corynebacterium dapA gene, the 917 bp Nco I to EcoR I
fragment of SEQ ID NO:3 inserted in pBT430 (pFS766, see
Example 1) was used.

For high level expression each of the plasmids was
transformed into E. coli strain BL21(DE3) [Studier
10   et al. (1986) J. Mol. Biol. 189:113-130]. Cultures were
grown in LB medium containing ampicillin (100 mg/L) at
25°C. At an optical density at 600 nm of approximately
1, IPTG (isopropylthio-β-galactoside, the inducer) was
added to a final concentration of 0.4 mM and incubation
15   was continued for 3 h at 25°C. The cells were collected
by centrifugation and resuspended in 1/20th (or 1/100th)
the original culture volume in 50 mM NaCl; 50 mM
Tris-Cl, pH 7.5; 1 mM EDTA, and frozen at -20°C. Frozen
aliquots of 1 mL were thawed at 37°C and sonicated, in
20   an ice-water bath, to lyse the cells. The lysate was
centrifuged at 4°C for 5 min at 15,000 rpm. The
supernatant was removed and the pellet was resuspended
in 1 mL of the above buffer.

The supernatant and pellet fractions of uninduced
25   and IPTG-induced cultures of BL21(DE3)/pBT442 or
BL21(DE3)/pFS766 were analyzed by SDS polyacrylamide gel
electrophoresis. The major protein visible by Coomassie
blue staining in the supernatant and pellet fractions of
both induced cultures had a molecular weight of
30   32-34 kd, the expected size for DHDPS. Even in the
uninduced cultures this protein was the most prominent
protein produced.

In the BL21(DE3)/pBT442 IPTG-induced culture about
80% of the DHDPS protein was in the supernatant and
35   DHDPS represented 10-20% of the total protein in the

extract. In the BL21(DE3)/pFS766 IPTG-induced culture
more than 50% of the DHDPS protein was in the pellet
fraction. The pellet fractions in both cases were
90-95% pure DHDPS, with no other single protein present

5   in significant amounts. Thus, these fractions were pure
enough for use in the generation of antibodies. The
pellet fractions containing 2-4 milligrams of either
E. coli DHDPS or Corynebacterium DHDPS were solubilized
in 50 mM NaCl; 50 mM Tris-Cl, pH 7.5; 1 mM EDTA, 0.2 mM

10   dithiothreitol, 0.2% SDS and sent to Hazelton Research
Facility (310 Swampridge Road, Denver, PA 17517) to have
rabbit antibodies raised against the proteins.

DHDPS enzyme activity was assayed as follows:
Assay mix (for 10 X 1.0 mL assay tubes or 40 X 0.25 mL

15   for microtiter dish); made fresh, just before use:

2.5mL	H_2O
0.5mL	1.0M Tris-HCl pH8.0
0.5mL	0.1M Na Pyruvate
0.5mL	o-Aminobenzaldehyde (10mg/mL in ethanol)
25μL	1.0M DL-Aspartic-β-semialdehyde (ASA) in 1.0N HCl

	Assay (1.0mL):	MicroAssay (0.25mL):
DHDPS assay mix	0.40mL	0.10mL
enzyme extract + H_2O;	0.10mL	.025mL
10mM L-lysine	5μL or 20μL	1μL or 5μL

Incubate at 30°C for desired time. Stop by addition of:

30   1.0N HCl   0.50mL   0.125mL

Color allowed to develop for 30-60 min. Precipitate
spun down in eppendorf centrifuge. OD_{540} vs 0 min read
as blank. For MicroAssay, aliquot 0.2 mL into

35   microtiter well and read at OD_{530}.

44

The specific activity of E. coli DHDPS in the
supernatant fraction of induced extracts was about
50 OD$_{540}$ units per minute per milligram protein in a
1.0 mL assay. E. coli DHDPS was sensitive to the
5   presence of L-lysine in the assay. Fifty percent
inhibition was found at a concentration of about 0.5 mM.
For Corynebacterium DHDPS, the activity was measured in
the supernatant fraction of uninduced extracts, rather
than induced extracts. Enzyme activity was about 4 OD$_{530}$
10 units per minute per milligram protein in a 0.25 mL
assay. In contrast to E. coli DHDPS, Corynebacterium
DHDPS was not inhibited at all by L-lysine, even at a
concentration of 70 mM.

<div align="center">EXAMPLE 3</div>

15   Isolation of the E. coli lysC Gene and mutations
            in lysC resulting in lysine-insensitive AKIII
        The E. coli lysC gene has been cloned, restriction
endonuclease mapped and sequenced previously [Cassan
et al. (1986) J. Biol. Chem. 261:1052-1057]. For the
20 present invention the lysC gene was obtained on a
bacteriophage lambda clone from an ordered library of
3400 overlapping segments of cloned E. coli DNA
constructed by Kohara, Akiyama and Isono [Kohara et al.
(1987) Cell 50:595-508]. This library provides a
25 physical map of the whole E. coli chromosome and ties
the physical map to the genetic map. From the knowledge
of the map position of lysC at 90 min. on the E. coli
genetic map [Theze et al. (1974) J. Bacteriol.
117:133-143], the restriction endonuclease map of the
30 cloned gene [Cassan et al. (1986) J. Biol. Chem.
261:1052-1057], and the restriction endonuclease map of
the cloned DNA fragments in the E. coli library [Kohara
et al. (1987) Cell 50:595-508], it was possible to
choose lambda phages 4E5 and 7A4 [Kohara et al. (1987)
35 Cell 50:595-508] as likely candidates for carrying the

45

lysC gene. The phages were grown in liquid culture from
single plaques as described [see Current Protocols in
Molecular Biology (1987) Ausubel et al. eds. John Wiley
& Sons New York] using LE392 as host [see Sambrook

5   et al. (1989) Molecular Cloning: a Laboratory Manual,
Cold Spring Harbor Laboratory Press]. Phage DNA was
prepared by phenol extraction as described [see Current
Protocols in Molecular Biology (1987) Ausubel et al.
eds. John Wiley & Sons, New York].

10   From the sequence of the gene several restriction
endonuclease fragments diagnostic for the lysC gene were
predicted, including an 1860 bp EcoR I-Nhe I fragment, a
2140 bp EcoR I-Xmn I fragment and a 1600 bp
EcoR I-BamH I fragment. Each of these fragments was

15 detected in both of the phage DNAs confirming that these
carried the lysC gene. The EcoR I-Nhe I fragment was
isolated and subcloned in plasmid pBR322 digested with
the same enzymes, yielding an ampicillin-resistant,
tetracycline-sensitive E. coli transformant. The

20 plasmid was designated pBT436.

        To establish that the cloned lysC gene was
functional, pBT436 was transformed into E. coli strain
Gif106M1 (E. coli Genetic Stock Center strain CGSC-5074)
which has mutations in each of the three E. coli AK

25 genes [Theze et al. (1974) J. Bacteriol. 117:133-143].
This strain lacks all AK activity and therefore requires
diaminopimelate (a precursor to lysine which is also
essential for cell wall biosynthesis), threonine and
methionine. In the transformed strain all these

30 nutritional requirements were relieved demonstrating
that the cloned lysC gene encoded functional AKIII.

        Addition of lysine (or diaminopimelate which is
readily converted to lysine in vivo) at a concentration
of approximately 0.2 mM to the growth medium inhibits

35 the growth of Gif106M1 transformed with pBT436. M9

46

media [see Sambrook et al. (1989) Molecular Cloning: a
Laboratory Manual, Cold Spring Harbor Laboratory Press]
supplemented with the arginine and isoleucine, required
for Gif106M1 growth, and ampicillin, to maintain

5   selection for the pBT436 plasmid, was used. This
inhibition is reversed by addition of threonine plus
methionine to the growth media. These results indicated
that AKIII could be inhibited by exogenously added
lysine leading to starvation for the other amino acids

10   derived from aspartate. This property of pBT436-
transformed Gif106M1 was used to select for mutations in
lysC that encoded lysine-insensitive AKIII.

Single colonies of Gif106M1 transformed with pBT436
were picked and resuspended in 200 μL of a mixture of

15   100 μL 1% lysine plus 100 μL of M9 media. The entire
cell suspension containing 10^7–10^8 cells was spread on a
petri dish containing M9 media supplemented with the
arginine, isoleucine, and ampicillin. Sixteen petri
dishes were thus prepared. From 1 to 20 colonies

20   appeared on 11 of the 16 petri dishes. One or two (if
available) colonies were picked and retested for lysine
resistance and from this nine lysine-resistant clones
were obtained. Plasmid DNA was prepared from eight of
these and re-transformed into Gif106M1 to determine

25   whether the lysine resistance determinant was plasmid-
borne. Six of the eight plasmid DNAs yielded lysine-
resistant colonies. Three of these six carried lysC
genes encoding AKIII that was uninhibited by 15mM
lysine, whereas wild type AKIII is 50% inhibited by

30   0.3–0.4 mM lysine and >90% inhibited by 1 mM lysine (see
Example 2 for details).

To determine the molecular basis for lysine-
resistance the sequences of the wild type lysC gene and
three mutant genes were determined. A method for "Using

35   mini-prep plasmid DNA for sequencing double stranded

47

templates with sequenase™" [Kraft et al. (1988)
BioTechniques 6:544-545] was used. Oligonucleotide
primers, based on the published lysC sequence and spaced
approximately every 200 bp, were synthesized to
5   facilitate the sequencing. The sequence of the wild
type lysC gene cloned in pBT436 (SEQ ID NO:5) differed
from the published lysC sequence in the coding region at
5 positions. Four of these nucleotide differences were
at the third position in a codon and would not result in
10   a change in the amino acid sequence of the AKIII
protein. One of the differences would result in a
cysteine to glycine substitution at amino acid 58 of
AKIII. These differences are probably due to the
different strains from which the lysC genes were cloned.
15   The sequences of the three mutant lysC genes that
encoded lysine-insensitive AK each differed from the
wild type sequence by a single nucleotide, resulting in
a single amino acid substitution in the protein. Mutant
M2 had an A substituted for a G at nucleotide 954 of
20   SEQ ID NO:5 resulting in an isoleucine for methionine
substitution at amino acid 318 and mutants M3 and M4 had
identical T for C substitutions at nucleotide 1055 of
SEQ ID NO:5 resulting in an isoleucine for threonine
substitution at amino acid 352. Thus, either of these
25   single amino acid substitutions is sufficient to render
the AKIII enzyme insensitive to lysine inhibition.
           An Nco I (CCATGG) site was inserted at the
translation initiation codon of the lysC gene using the
following oligonucleotides:
30

SEQ ID NO:6:
GATCCATGGC TGAAATTGTT GTCTCCAAAT TTGGCG

SEQ ID NO:7:
35   GTACCGCCAA ATTTGGAGAC AACAATTTCA GCCATG

48

When annealled these oligonucleotides have BamH I and
Asp 718 "sticky" ends. The plasmid pBT436 was digested
with BamH I, which cuts upstream of the lysC coding
sequence and Asp 718 which cuts 31 nucleotides
5   downstream of the initiation codon. The annealled
oligonucleotides were ligated to the plasmid vector and
E. coli transformants were obtained. Plasmid DNA was
prepared and screened for insertion of the
oligonucleotides based on the presence of an Nco I site.
10   A plasmid containing the site was sequenced to assure
that the insertion was correct, and was designated
pBT457. In addition to creating an Nco I site at the
initiation codon of lysC, this oligonucleotide insertion
changed the second codon from TCT, coding for serine, to
15   GCT, coding for alanine. This amino acid substitution
has no apparent effect on the AKIII enzyme activity.

The lysC gene was cut out of plasmid pBT457 as a
1560 bp Nco I-EcoR I fragment and inserted into the
expression vector pBT430 digested with the same enzymes,
20   yielding plasmid pBT461. For expression of the mutant
lysC-M4 gene pBT461 was digested with Kpn I-EcoR I,
which removes the wild type lysC gene from about 30
nucleotides downstream from the translation start codon,
and inserting the analogous Kpn I-EcoR I fragments from
25   the mutant genes yielding plasmid pBT492.

EXAMPLE 4

Construction of Chimeric dapA

Genes for Expression in the Seeds of Plants

A seed-specific expression cassette (Figure 4) is
30   composed of the promoter and transcription terminator
from the gene encoding the β subunit of the seed storage
protein phaseolin from the bean Phaseolus vulgaris
[Doyle et al. (1986) J. Biol. Chem. 261:9228-9238]. The
phaseolin cassette includes about 500 nucleotides
35   upstream (5') from the translation initiation codon and

49

about 1650 nucleotides downstream (3') from the
translation stop codon of phaseolin. Between the 5' and
3' regions are the unique restriction endonuclease sites
Nco I (which includes the ATG translation initiation
5   codon), Sma I, Kpn I and Xba I. The entire cassette is
flanked by Hind III sites.

Plant amino acid biosynthetic enzymes are known to
be localized in the chloroplasts and therefore are
synthesized with a chloroplast targeting signal.
10 Bacterial proteins such as DHDPS and AKIII have no such
signal. A chloroplast transit sequence (cts) was
therefore fused to the dapA and lysC-M4 coding sequence
in the chimeric genes. The cts used was based on the
the cts of the small subunit of ribulose 1,5-bisphos-
15 phate carboxylase from soybean [Berry-Lowe et al. (1982)
J. Mol. Appl. Genet. 1:483-498]. The oligonucleotides
SEQ ID NOS:8-11 were synthesized and used as described
below.

Three chimeric genes were created:
20   No. 1)   phaseolin 5' region/cts/lysC-M4/phaseolin
                    3' region
           No. 2)   phaseolin 5' region/cts/ecodapA/phaseolin
                    3' region
           No. 3)   phaseolin 5' region/cts/cordapA/phaseolin
25   3' region

Oligonucleotides SEQ ID NO:8 and SEQ ID NO:9, which
encode the carboxy terminal part of the chloroplast
targeting signal, were annealed, resulting in Nco I
compatible ends, purified via polyacrylamide gel
30 electrophoresis, and inserted into Nco I digested
pBT461. The insertion of the correct sequence in the
correct orientation was verified by DNA sequencing
yielding pBT496. Oligonucleotides SEQ ID NO:10 and SEQ
ID NO:11, which encode the amino terminal part of the
35 chloroplast targeting signal, were annealed, resulting

in Nco I compatible ends, purified via polyacrylamide
gel electrophoresis, and inserted into Nco I digested
pBT496. The insertion of the correct sequence in the
correct orientation was verified by DNA sequencing
5 yielding pBT521. Thus the cts was fused to the lysC
gene.

   To fuse the cts to the lysC-M4 gene, pBT521 was
digested with Sal I, and an approximately 900 bp DNA
fragment that included the cts and the amino terminal
10 coding region of lysC was isolated. This fragment was
inserted into Sal I digested pBT492, effectively
replacing the amino terminal coding region of lysC-M4
with the fused cts and the amino terminal coding region
of lysC. Since the mutation that resulted in lysine-
15 insensitivity was not in the replaced fragment, the new
plasmid, pBT523, carried the cts fused to lysC-M4.

   The 1600 bp Nco I-Hpa I fragment containing the cts
fused to lysC-M4 plus about 90 bp of 3' non-coding
sequence was isolated and inserted into the seed-
20 specific expression cassette digested with Nco I and
Sma I (chimeric gene No. 1), yielding plasmid pBT544.

   Before insertion into the expression cassette, the
ecodapA gene was modified to insert a restriction
endonuclease site, Kpn I, just after the translation
25 stop codon. The oligonucleotides SEQ ID NOS:12-13 were
synthesized for this purpose:

SEQ ID NO:12:
CCGGTTTGCT GTAATAGGTA CCA
30
SEQ ID NO:13:
AGCTTGGTAC CTATTACAGC AAACCGGCAT G

   Oligonucleotides SEQ ID NO:12 and SEQ ID NO:13 were
35 annealed, resulting in an Sph I compatible end on one
end and a Hind III compatible end on the other and

inserted into Sph I plus Hind III digested pBT437. The
insertion of the correct sequence was verified by DNA
sequencing yielding pBT443.

An 880 bp Nco I-Kpn I fragment from pBT443
.5   containing the entire ecodapA coding region was isolated
from an agarose gel following electrophoresis and
inserted into the seed-specific expression cassette
digested with Nco I and Kpn I, yielding plasmid pBT494.
Oligonucleotides SEQ ID NO:8-11 were used as described
10   above to add a cts to the ecodapA coding region in the
seed-specific expression cassette, yielding chimeric
gene No. 2 in pBT520.

An 870 bp Nco I-EcoR I fragment from pFS766
containing the entire cordapA coding region was isolated
15   from an agarose gel following electrophoresis and
inserted into the leaf expression cassette digested with
Nco I and EcoR I, yielding plasmid pFS789 To attach
the cts to the cordapA gene a DNA fragment containing
the entire cts was prepared using PCR. The template DNA
20   was pBT544 and the oligonucleotide primers used were:

SEQ ID NO:14:
GCTTCCTCAA TGATCTCCTC CCCAGCT

25   SEQ ID NO:15:
CATTGTACTC TTCCACCGTT GCTAGCAA

PCR was performed using a Perkin-Elmer Cetus kit
according to the instructions of the vendor on a
30   thermocycler manufactured by the same company. The
PCR-generated 160 bp fragment was treated with T4 DNA
polymerase in the presence of the 4 deoxyribonucleotide
triphosphates to obtain a blunt-ended fragment. The cts
fragment was inserted into the Nco I containing the
35   start codon of the cordapA gene which had been digested
and treated with the Klenow fragment of DNA polymerase

52

to fill in the 5' overhangs. The inserted fragment and
the vector/insert junctions were determined to be
correct by DNA sequencing.

A 1030 bp Nco I-Kpn I fragment containing the cts
5   attached to the cordapA coding region was isolated from
an agarose gel following electrophoresis and inserted
into the phaseolin seed expression cassette digested
with Nco I and Kpn I, yielding plasmid pFS889 containing
chimeric gene No. 3.

10   EXAMPLE 5
                  Transformation of Rapeseed with the
                    Phaseolin Promoter/cts/cordapA and
              Phaseolin Promoter/cts/lysC-M4 Chimeric Genes

The chimeric gene cassettes, phaseolin 5' region/
15   cts/cordapA/phaseolin 3' region, phaseolin 5' region/
cts/lysC-M4/phaseolin 3', and phaseolin 5' region/
cts/cordapA/phaseolin 3' region plus phaseolin 5'
region/cts/lysC-M4/phaseolin 3' (Example 4) were
inserted into the binary vector pZS199 (Figure 5A). In
20   pZS199 the 35S promoter from Cauliflower Mosaic Virus
drives expression of the NPT II.

The phaseolin 5' region/cts/cordapA/phaseolin 3'
region chimeric gene cassette was modified using
oligonucleotide adaptors to convert the Hind III sites
25   at each end to BamH I sites. The gene cassette was then
isolated as a 2.7 kb BamH I fragment and inserted into
BamH I digested pZS199, yielding plasmid pFS926
(Figure 5B). This binary vector has the chimeric gene,
phaseolin 5' region/cts/cordapA/phaseolin 3' region
30   inserted in the same orientation as the 35S/NPT II/nos
3' marker gene.

To insert the phaseolin 5' region/cts/lysC-
M4/phaseolin 3' region, the gene cassette was isolated
as a 3.3 kb EcoR I to Spe I fragment and inserted into
35   EcoR I plus Xba I digested pZS199, yielding plasmid

pBT593. This binary vector has the chimeric gene,
phaseolin 5' region/cts/lysC-M4/phaseolin 3' region
inserted in the same orientation as the 35S/NPT II/nos
3' marker gene.

5   To combine the two cassettes, the EcoR I site of
pBT593 was converted to a BamH I site using
oligonucleotide adaptors, the resulting vector was cut
with BamH I and the phaseolin 5' region/cts/cordapA/
phaseolin 3' region gene cassette was isolated as a

10   2.7 kb BamH I fragment and inserted, yielding pBT597.
This binary vector has both chimeric genes, phaseolin 5'
region/cts/cordapA/phaseolin 3' region and phaseolin 5'
region/cts/lysC-M4/phaseolin 3' region inserted in the
same orientation as the 35S/NPT II/nos 3' marker gene.

15   Brassica napus cultivar "Westar" was transformed by
co-cultivation of seedling pieces with disarmed
Agrobacterium tumefaciens strain LBA4404 carrying the
the appropriate binary vector.

          B. napus seeds were sterilized by stirring in 10%

20   Chlorox, 0.1% SDS for thirty min, and then rinsed
thoroughly with sterile distilled water. The seeds were
germinated on sterile medium containing 30 mM CaCl2 and
1.5% agar, and grown for six d in the dark at 24°C.

          Liquid cultures of Agrobacterium for plant

25   transformation were grown overnight at 28°C in Minimal A
medium containing 100 mg/L kanamycin. The bacterial
cells were pelleted by centrifugation and resuspended at
a concentration of 10^8 cells/mL in liquid Murashige and
Skoog Minimal Organic medium containing 100 uM

30   acetosyringone.

          B. napus seedling hypocotyls were cut into 5 mm
segments which were immediately placed into the
bacterial suspension. After 30 min, the hypocotyl
pieces were removed from the bacterial suspension and

35   placed onto BC-35 callus medium containing 100 uM

54

acetosyringone. The plant tissue and _Agrobacteria_ were
co-cultivated for three d at 24°C in dim light.

The co-cultivation was terminated by transferring
the hypocotyl pieces to BC-35 callus medium containing
5   200 mg/L carbenicillin to kill the _Agrobacteria_, and
25 mg/L kanamycin to select for transformed plant cell
growth. The seedling pieces were incubated on this
medium for three weeks at 24°C under continuous light.

After three weeks, the segments were transferred to
10   BS-48 regeneration medium containing 200 mg/L
carbenicillin and 25 mg/L kanamycin. Plant tissue was
subcultured every two weeks onto fresh selective
regeneration medium, under the same culture conditions
described for the callus medium. Putatively transformed
15   calli grew rapidly on regeneration medium; as calli
reached a diameter of about 2 mm, they were removed from
the hypocotyl pieces and placed on the same medium
lacking kanamycin

Shoots began to appear within several weeks after
20   transfer to BS-48 regeneration medium. As soon as the
shoots formed discernable stems, they were excised from
the calli, transferred to MSV-1A elongation medium, and
moved to a 16:8-h photoperiod at 24°C.

Once shoots had elongated several internodes, they
25   were cut above the agar surface and the cut ends were
dipped in Rootone. Treated shoots were planted directly
into wet Metro-Mix 350 soiless potting medium. The pots
were covered with plastic bags which were removed when
the plants were clearly growing, after about ten d.
30   Results of the transformation are shown in Table 1.
Transformed plants were obtained with each of the binary
vectors.

55

Minimal A Bacterial Growth Medium

Dissolve in distilled water:

      10.5 g potassium phosphate, dibasic

      4.5 g potassium phosphate, monobasic

5   1.0 g ammonium sulfate

      0.5 g sodium citrate, dihydrate

Make up to 979 mL with distilled water

Autoclave

Add 20 mL filter-sterilized 10% sucrose

10 Add 1 mL filter-sterilized 1 M MgSO4

Brassica Callus Medium BC-35

Per liter:

      Murashige and Skoog Minimal Organic Medium

15   (MS salts, 100 mg/L i-inositol, 0.4 mg/L thiamine; GIBCO
#510-3118)

      30 g sucrose

      18 g mannitol

      0.5 mg/L 2,4-D

20   0.3 mg/L kinetin

      0.6% agarose

      pH 5.8

Brassica Regeneration Medium BS-48

25   Murashige and Skoog Minimal Organic Medium

      Gamborg B5 Vitamins (SIGMA #1019)

      10 g glucose

      250 mg xylose

      600 mg MES

30   0.4% agarose

      pH 5.7

Filter-sterilize and add after autoclaving:

      2.0 mg/L zeatin

      0.1 mg/L IAA

35

56

<u>Brassica Shoot Elongation Medium MSV-1A</u>

Murashige and Skoog Minimal Organic Medium

Gamborg B5 Vitamins

10 g sucrose

5   0.6% agarose

pH 5.8

<u>TABLE 1</u>
Canola transformants

BINARY VECTOR	NUMBER OF CUT ENDS	NUMBER OF KAN^R CALLI	NUMBER OF SHOOTING CALLI	NUMBER OF PLANTS
pZS199	120	41	5	2
pFS926	600	278	52	28
pBT593	600	70	10	3
pBT597	600	223	40	23

Plants were grown under a 16:8-h photoperiod, with
10   a daytime temperature of 23°C and a nightime temperature
of 17°C. When the primary flowering stem began to
elongate, it was covered with a mesh pollen-containment
bag to prevent outcrossing. Self-pollination was
facilitated by shaking the plants several times each
15   day. Mature seeds derived from self-pollinations were
harvested about three months after planting.

A partially defatted seed meal was prepared as
follows: 40 milligrams of mature dry seed was ground
with a mortar and pestle under liquid nitrogen to a fine
20   powder. One milliliter of hexane was added and the
mixture was shaken at room temperature for 15 min. The
meal was pelleted in an eppendorf centrifuge, the hexane
was removed and the hexane extraction was repeated.
Then the meal was dried at 65° for 10 min until the
25   hexane was completely evaporated leaving a dry powder.
Total proteins were extracted from mature seeds as
follows. Approximately 30-40 mg of seeds were put into
a 1.5 mL disposable plastic microfuge tube and ground in

57

0.25 mL of 50 mM Tris-HCl pH 6.8, 2 mM EDTA, 1% SDS, 1%
β-mercaptoethanol. The grinding was done using a
motorized grinder with disposable plastic shafts
designed to fit into the microfuge tube. The resultant
5   suspensions were centrifuged for 5 min at room
temperature in a microfuge to remove particulates.
Three volumes of extract was mixed with 1 volume of 4 X
SDS-gel sample buffer (0.17M Tris-HCl pH6.8, 6.7% SDS,
16.7% β-mercaptoethanol, 33% glycerol) and 5 μL from
10   each extract were run per lane on an SDS polyacrylamide
gel, with bacterially produced DHDPS or AKIII serving as
a size standard and protein extracted from untransformed
tobacco seeds serving as a negative control. The
proteins were then electrophoretically blotted onto a
15   nitrocellulose membrane. The membranes were exposed to
the DHDPS or AKIII antibodies at a 1:5000 dilution of
the rabbit serum using standard protocol provided by
BioRad with their Immun-Blot Kit. Following rinsing to
remove unbound primary antibody the membranes were
20   exposed to the secondary antibody, donkey anti-rabbit Ig
conjugated to horseradish peroxidase (Amersham) at a
1:3000 dilution. Following rinsing to remove unbound
secondary antibody, the membranes were exposed to
Amersham chemiluminescence reagent and X-ray film.
25   Eight of eight FS926 transformants and seven of
seven BT597 transformants expressed the DHDPS protein.
The single BT593 transformant and five of seven BT597
transformants expressed the AKIII-M4 protein (Table 2).
         To measure free amino acid composition of the
30   seeds, free amino acids were extracted from 40
milligrams of the defatted meal in 0.6 mL of
methanol/chloroform/water mixed in ratio of 12v/5v/3v
(MCW) at room temperature. The mixture was vortexed and
then centrifuged in an eppendorf microcentrifuge for
35   about 3 min. Approximately 0.6 mL of supernatant was

58

decanted and an additional 0.2 mL of MCW was added to
the pellet which was then vortexed and centrifuged as
above. The second supernatant, about 0.2 mL, was added
to the first. To this, 0.2 mL of chloroform was added
5   followed by 0.3 mL of water. The mixture was vortexed
and then centrifuged in an eppendorf microcentrifuge for
about 3 min, the upper aqueous phase, approximately
1.0 mL, was removed, and was dried down in a Savant
Speed Vac Concentrator. The samples were hydrolyzed in
10 6N hydrochloric acid, 0.4% β-mercaptoethanol under
nitrogen for 24 h at 110-120°C; 1/4 of the sample was
run on a Beckman Model 6300 amino acid analyzer using
post-column ninhydrin detection. Relative free amino
acid levels in the seeds were compared as ratios of
15 lysine or threonine to leucine, thus using leucine as an
internal standard.
         There was a good correlation between transformants
expressing higher levels of DHDPS protein and those
having higher levels of free lysine. The highest
20 expressing lines showed a greater than 100-fold increase
in free lysine level in the seeds. There has been no
greater accumulation of free lysine due to expression of
AKIII-M4 along with _Corynebacteria_ DHDPS compared to
expression of _Corynebacteria_ DHDPS alone. The
25 transformant that expressed AKIII-M4 in the absence of
Corynebacteria DHDPS showed a 5-fold increase in the
level of free threonine in the seeds. A high level of
α-aminoadipic acid, indicative of lysine catabolism, was
observed in many of the transformed lines. Thus,
30 prevention of lysine catabolism by inactivation of
lysine ketoglutarate reductase should further increase
the accumulation of free lysine in the seeds.
Alternatively, incorporation of lysine into a peptide or
lysine-rich protein would prevent catabolism and lead to
35 an increase in the accumulation of lysine in the seeds.

To measure the total amino acid composition of
mature seeds, 2 milligrams of the defatted meal were
hydrolyzed in 6N hydrochloric acid, 0.4% β-mercapto-
ethanol under nitrogen for 24 h at 110-120°C; 1/100 of
5   the sample was run on a Beckman Model 6300 amino acid
analyzer using post-column ninhydrin detection.
Relative amino acid levels in the seeds were compared as
percentages of lysine, threonine or α-aminoadipic acid
to total amino acids. There was a good correlation
10   between transformants expressing DHDPS protein and those
having high levels of lysine. Seeds with a 5-100%
increase in the lysine level, compared to the
untransformed control, were observed. In the seeds with
the highest levels, lysine makes up 11-13% of the total
15   seed amino acids, considerably higher than any
previously known rapeseed seed.

TABLE 2

FS926 Transformants: phaseolin 5' region/cts/cor_dapA_/phaseolin 3'
BT593 Transformants: phaseolin 5' region/cts/_lysC_-M4/phaseolin 3'
BT597 Transformants: phaseolin 5' region/cts/_lysC_-M4/phaseolin 3'
                       phaseolin 5' region/cts/cor_dapA_/phaseolin 3'

LINE	FREE AMINO ACIDS K/L	T/L	AA/L	WESTERN CORYNE. DHDPS	WESTERN E. COLI AKIII-M4	% TOTAL AMINO ACIDS K	T	AA
WESTAR	0.8	2.0	0	-	-	6.5	5.6	0
ZS199	1.3	3.2	0	-	-	6.3	5.4	0
FS926-3	140	2.0	16	++++	-	12	5.1	1.0
FS926-9	110	1.7	12	++++	-	11	5.0	0.8
FS926-11	7.9	2.0	5.2	++	-	7.7	5.2	0
FS926-6	14	1.8	4.6	+++	-	8.2	5.9	0
FS926-22	3.1	1.3	0.3	+	-	6.9	5.7	0
FS926-27	4.2	1.9	1.1	++	-	7.1	5.6	0
FS926-29	38	1.8	4.7	++++	-	12	5.2	1.6
FS926-68	4.2	1.8	0.9	++	-	8.3	5.5	0
BT593-42	1.4	11	0	-	++	6.3	6.0	0
BT597-14	6.0	2.6	4.3	++	+/-	7.0	5.3	0

BT597-145	1.3	2.9	0	+	-			
BT597-4	38	3.7	4.5	++++	++++	13	5.6	1.6
BT597-68	4.7	2.7	1.5	++	+	6.9	5.8	0
BT597-100	9.1	1.9	1.7	+++	++	6.6	5.7	0
BT597-148	7.6	2.3	0.9	+++	+	7.3	5.7	0
BT597-169	5.6	2.6	1.7	+++	+++	6.6	5.7	0

AA is α-amino adipic acid

EXAMPLE 6

Transformation of Soybean with the
Phaseolin Promoter/cts/cordapA and
Phaseolin Promoter/cts/lysC-M4 Chimeric Genes

The chimeric gene cassettes, phaseolin 5' region/
cts/cordapA/phaseolin 3' region plus phaseolin 5'
region/cts/lysC-M4/phaseolin 3', (Example 4) were
inserted into the soybean transformation vector pBT603
(Figure 6A). This vector has a soybean transformation
marker gene consisting of the 35S promoter from
Cauliflower Mosaic Virus driving expression of the
E. coli β-glucuronidase (GUS) gene [Jefferson et al.
(1986) Proc. Natl. Acad. Sci. USA 83:8447-8451] with the
Nos 3' region in a modified pGEM9Z plasmid.

To insert the phaseolin 5' region/cts/lysC-M4/
phaseolin 3' region, the gene cassette was isolated as a
3.3 kb Hind III fragment and inserted into Hind III
digested pBT603, yielding plasmid pBT609. This vector
has the chimeric gene, phaseolin 5' region/
cts/lysC-M4/phaseolin 3' region inserted in the opposite
orientation from the 35S/GUS/Nos 3' marker gene.

The phaseolin 5' region/cts/cordapA/phaseolin
3'region chimeric gene cassette was modified using
oligonucleotide adaptors to convert the Hind III sites
at each end to BamH I sites. The gene cassette was then
isolated as a 2.7 kb BamH I fragment and inserted into
BamH I digested pBT609, yielding plasmid pBT614

61

(Figure 6B). This vector has both chimeric genes,
phaseolin 5' region/cts/cordapA/phaseolin 3' region plus
phaseolin 5' region/cts/lysC-M4/phaseolin 3' inserted in
the same orientation, and both are in the opposite
5   orientation from the 35S/GUS/Nos 3' marker gene.

Plasmid pBT614 was introduced into soybean via
transformation by Agracetus Company (Middleton, WI),
according to the procedure described in United States
Patent No. 5,015,580. Seeds from five transformed lines
10   were obtained and analyzed.

It was expected that the transgenes would be
segregating in the R1 seeds of the transformed plants.
To identify seeds that carried the transformation marker
gene, a small chip of the seed was cut off with a razor
15   and put into a well in a a disposable plastic microtiter
plate. A GUS assay mix consisting of 100 mM NaH_2PO_4,
10 mM EDTA, 0.5 mM $K_4Fe(CN)_6$, 0.1% Triton X-100,
0.5 mg/mL 5-Bromo-4-chloro-3-indolyl β-D-glucuronic acid
was prepared and 0.15 mL was added to each microtiter
20   well. The microtiter plate was incubated at 37° for
45 minutes. The development of blue color indicated the
expression of GUS in the seed.

Four of five transformed lines showed approximately
3:1 segregation for GUS expression (Table 3). This
25   indicates that the GUS gene was inserted at a single
site in the soybean genome. The other transformant
showed 9:1 segregation, suggesting that the GUS gene was
inserted at two sites.

A meal was prepared from a fragment of individual
30   seeds by grinding into a fine powder. Total proteins
were extracted from the meal by adding 1 mg to 0.1 mL of
43 mM Tris-HCl pH 6.8, 1.7% SDS, 4.2% β-mercaptoethanol,
8% glycerol, vortexing the suspension, boiling for 2-3
minutes and vortexing again. The resultant suspensions
35   were centrifuged for 5 min at room temperature in a

62

microfuge to remove particulates and 10 µL from each
extract were run per lane on an SDS polyacrylamide gel,
with bacterially produced DHDPS or AKIII serving as a
size standard. The proteins were then electro-
5 phoretically blotted onto a nitrocellulose membrane.
The membranes were exposed to the DHDPS or AKIII
antibodies, at a 1:5000 or 1:1000 dilution,
respectively, of the rabbit serum using standard
protocol provided by BioRad with their Immun-Blot Kit.
10 Following rinsing to remove unbound primary antibody the
membranes were exposed to the secondary antibody, donkey
anti-rabbit Ig conjugated to horseradish peroxidase
(Amersham) at a 1:3000 dilution. Following rinsing to
remove unbound secondary antibody, the membranes were
15 exposed to Amersham chemiluminescence reagent and X-ray
film.
        Four of five transformants expressed the DHDPS
protein. In the four transformants that expressed
DHDPS, there was excellent correlation between
20 expression of GUS and DHDPS in individual seeds
(Table 3). Therefore, the GUS and DHDPS genes are
integrated at the same site in the soybean genome. Two
of five transformants expressed the AKIII protein, and
again there was excellent correlation between expression
25 of AKIII, GUS and DHDPS in individual seeds (Table 3).
Thus, in these two transformants the GUS, AKIII and
DHDPS genes are integrated at the same site in the
soybean genome. One transformant expressed only GUS in
its seeds.
30   To measure free amino acid composition of the
seeds, free amino acids were extracted from 8-10
milligrams of the meal in 1.0 mL of methanol/chloro-
form/water mixed in ratio of 12v/5v/3v (MCW) at room
temperature. The mixture was vortexed and then
35 centrifuged in an eppendorf microcentrifuge for about

3 min; approximately 0.8 mL of supernatant was decanted.
To this supernatant, 0.2 mL of chloroform was added
followed by 0.3 mL of water. The mixture was vortexed
and then centrifuged in an eppendorf microcentrifuge for
5   about 3 min, the upper aqueous phase, approximately
1.0 mL, was removed, and was dried down in a Savant
Speed Vac Concentrator. The samples were hydrolyzed in
6N hydrochloric acid, 0.4% β-mercaptoethanol under
nitrogen for 24 h at 110-120°C; 1/10 of the sample was
10   run on a Beckman Model 6300 amino acid analyzer using
post-column ninhydrin detection. Relative free amino
acid levels in the seeds were compared as ratios of
lysine to leucine, thus using leucine as an internal
standard.
15   There was excellent correlation between
transformants expressing Corynebacteria DHDPS protein
and those having higher levels of free lysine. From
20 fold to 120-fold increases in free lysine level was
observed in seeds expressing Corynebacteria DHDPS. A
20   high level of saccharopine, indicative of lysine
catabolism, was observed in seeds the contained high
levels of lysine.
         To measure the total amino acid composition of
mature seeds, 1-1.4 milligrams of the seed meal was
25   hydrolyzed in 6N hydrochloric acid, 0.4% β-mercapto-
ethanol under nitrogen for 24 h at 110-120°C; 1/50 of
the sample was run on a Beckman Model 6300 amino acid
analyzer using post-column ninhydrin detection. Lysine
(and other amino acid) levels in the seeds were compared
30   as percentages of the total amino acids.
         There was excellent correlation between seeds
expressing Corynebacteria DHDPS protein and those having
high levels of lysine. Seeds with a 5-35% increase in
the lysine level, compared to the untransformed control,
35   were observed. In these seeds lysine makes up 7.5-7.7%

64

of the total seed amino acids, considerably higher than
any previously known soybean seed.

TABLE 3

LINE-SEED	GUS	Free LYS/LEU	DHDPS	AKIII	% LYS TOT
A2396-145-4	-	0.9	-	-	5.75
A2396-145-8	-	1	-	-	
A2396-145-5	-	0.8			5.85
A2396-145-3	-	1			
A2396-145-9	+	2			
A2396-145-6	+	4.6			
A2396-145-1	+	8.7			
A2396-145-10	+	18.4			7.54
A2396-145-7	+	21.7	+	-	6.68
A2396-145-2	+	45.5	+	-	7.19
A5403-175-9	-	1.3			
A5403-175-4	-	1.2	-	-	6.01
A5403-175-3	-	1	-	-	6.02
A5403-175-7	+	1.5			
A5403-175-5	+	1.8			
A5403-175-1	+	6.2			
A5403-175-2	+	6.5			6.3
A5403-175-6	+	14.4			
A5403-175-8	+	47.8	+	-	7.67
A5403-175-10	+	124.3	+	-	7.49
A5403-181-9	+	1.4			
A5403-181-10	+	1.4	-	-	5.68
A5403-181-8	+	0.9			
A5403-181-6	+	1.5			
A5403-181-4	-	0.7	-	-	5.85
A5403-181-5	+	1.1			
A5403-181-2	-	1.8	-	-	5.59
A5403-181-3	+	2.7	-	-	5.5

65

A5403-181-7	+	1.9			
A5403-181-1	−	2.3			
A5403-183-9	−	0.8			
A5403-183-6	−	0.7	−	−	6.03
A5403-183-8	−	1.3			
A5403-183-4	−	1.3	−	−	6.04
A5403-183-5	+	0.9			
A5403-183-3	+	3.1			
A5403-183-1	+	3.3			
A5403-183-7	+	9.9			
A5403-183-10	+	22.3	+	+	6.74
A5403-183-2	+	23.1	+	+	7.3
A5403-196-8	−	0.9	−	−	5.92
A5403-196-6	+	8.3			
A5403-196-1	+	16.1	+	+	6.83
A5403-196-7	+	27.9			
A5403-196-3	+	52.8			
A5403-196-5	+	26			
A5403-196-2	+	16.2	+	+	
A5403-196-10	+	29	+	+	7.53
A5403-196-4	+	58.2	+	+	7.57
A5403-196-9	+	47.1			
wild type control	−	0.9	−	−	5.63

Eighteen additional transformed soybean lines were
obtained. Single seeds from the lines were analyzed for
GUS activity as described above, and all lines exhibited
5 GUS-positive seeds. Meal was prepared from single
seeds, or in some cases a pool of several seeds, and
assayed for expression of DHDPS and AKIII proteins via
western blot. Seventeen of the eighteen lines expressed
DHDPS, and fifteen of the eighteen expressed AKIII.
10 Again there was excellent correlation between seeds

66

expressing GUS, DHDPS and AKIII, indicating that the
genes are linked in the transformed lines.

The amino acid composition of the seeds from these
lines was determined as described above. Again seeds

5 expressing _Corynebacteria_ DHDPS protein showed increased
levels of lysine. Expression of DHDPS alone resulted in
5% to 40% increases in total seed lysine. Expression of
DHDPS along with AKIII-M4 results in lysine increases of
more than 400%. A summary of all the different

10 transformed lines is shown in Table 3A

TABLE 3A

LINE-SEED	GUS + to -	DHDPS	AKIII	% LYS TOT
A2396-145	6 to 4	+	-	7.5
A2396-233	3 to 1	+	+	25
A2396-234	15 to 1	+	+	16
A2396-248	4 to 10	+	-	6.3
A2396-263	14 to 2	-	-	
A2396-240	7 to 1	+	+	11
A2396-267	2 to 53	+	+	8.9
A2242-273	11 to 5	+	+	13
A2242-315	6 to 2	+	+	16
A2242-316	1 to 15	+	+	12
A5403-175	7 to 3	+	+	7.6
A5403-181	7 to 3	-	-	5.7
A5403-183	6 to 4	+	+	6
A5403-185	9 to 11	+	+	7.6 P
A5403-196	9 to 1	+	+	7.6
A5403-203	6 to 36	+	+	6.1 P
A5403-204	17 to 3	+	+	8.8 P
A5403-212	13 to 5	+	+	9.4 P
A5403-214	21 to 16	+	+	32
A5403-216	14 to 4	+	-	8.2 P

A5403-218	13 to 9	+	+	9.8 P
A5403-222	12 to 27	+	+	15
A5403-225	14 to 12	+	+	13

P indicates seeds were pooled before meal extraction and assay

EXAMPLE 7

Isolation of a Plant

Lysine Ketoglutarate Reductase Gene

        Lysine Ketoglutarate Reductase (LKR) enzyme activity has been observed in immature endosperm of developing maize seeds [Arruda et al. (1982) Plant Physiol. 69:988-989]. LKR activity increases sharply from the onset of endosperm development, reaches a peak level at about 20 d after pollination, and then declines [Arruda et al. (1983) Phytochemistry 22:2687-2689].

        In order to clone the corn LKR gene, RNA was isolated from developing seeds 19 d after pollination. This RNA was sent to Clontech Laboratories, Inc., (Palo Alto, CA) for the custom synthesis of a cDNA library in the vector Lambda Zap II. The conversion of the Lambda Zap II library into a phagemid library, then into a plasmid library was accomplished following the protocol provided by Clontech. Once converted into a plasmid library the ampicillin-resistant clones obtained carry the cDNA insert in the vector pBluescript SK(-). Expression of the cDNA is under control of the lacZ promoter on the vector.

        Two phagemid libraries were generated using the mixtures of the Lambda Zap II phage and the filamentous helper phage of 100 μL to 1 μL. Two additional libraries were generated using mixtures of 100 μL Lambda Zap II to 10 μL helper phage and 20 μL Lambda Zap II to 10 μL helper phage. The titers of the phagemid preparations were similar regardless of the mixture used and were about 2×10^3 ampicillin-resistant-

68

transfectants per mL with E. coli strain XL1-Blue as
the host and about 1 x 10^3 with DE126 (see below) as
host.

To select clones that carried the LKR gene a
5   specially designed E. coli host, DE126 was constructed.
Construction of DE126 occurred in several stages.
(1) A generalized transducing stock of coliphage P1vir
was produced by infection of a culture of TST1 [F$^-$,
araD139, delta(argF-lac)205, flb5301, ptsF25, relA1,
10 rpsL150, malE52::Tn10, deoC1, λ$^-$] (E. coli Genetic Stock
Center #6137) using a standard method (for Methods see
J. Miller, Experiments in Molecular Genetics).
(2) This phage stock was used as a donor in a
transductional cross (for Method see J. Miller,
15 Experiments in Molecular Genetics) with strain GIF106M1
[F$^-$, arg-, ilvA296, lysC1001, thrA1101, metL1000, λ$^-$,
rpsL9, malT1, xyl-7, mtl-2, thil(?), supE44(?)] (E. coli
Genetic Stock Center #5074) as the recipient.
Recombinants were selected on rich medium [L
20 supplemented with DAP] containing the antibiotic
tetracycline. The transposon Tn10, conferring
tetracycline resistance, is inserted in the malE gene of
strain TST1. Tetracycline-resistant transductants
derived from this cross are likely to contain up to
25 2 min of the E. coli chromosome in the vicinity of malE.
The genes malE and lysC are separated by less than
0.5 minutes, well within cotransduction distance.
(3) 200 tetracycline-resistant transductants were
thoroughly phenotyped; appropriate fermentation and
30 nutritional traits were scored. The recipient strain
GIF106M1 is completely devoid of aspartokinase isozymes
due to mutations in thrA, metL and lysC, and therefore
requires the presence of threonine, methionine, lysine
and meso-diaminopimelic acid (DAP) for growth.
35 Transductants that had inherited lysC$^+$ with malE::Tn10

from TST1 would be expected to grow on a minimal medium
that contains vitamin B1, L-arginine, L-isoleucine and
L-valine in addition to glucose which serves as a carbon
and energy source. Moreover strains having the genetic
5   constitution of lysC+, metL- and thrA- will only express
the lysine sensitive aspartokinase. Hence addition of
lysine to the minimal medium should prevent the growth
of the lysC+ recombinant by leading to starvation for
threonine, methionine and DAP. Of the 200 tetracycline
10   resistant transductants examined, 49 grew on the minimal
medium devoid of threonine, methionine and DAP.
Moreover, all 49 were inhibited by the addition of
L-lysine to the minimal medium. One of these
transductants was designated DE125. DE125 has the
15   phenotype of tetracycline resistance, growth
requirements for arginine, isoleucine and valine, and
sensitivity to lysine. The genotype of this strain is
F- malE52::Tn10 arg- ilvA296 thrA1101 metL1000 lambda-
rpsL9 malT1 xyl-7 mtl-2 thiI(?) supE44(?).

20   (4) This step involves production of a male
derivative of strain DE125. Strain DE125 was mated with
the male strain AB1528 [F'16/delta(gpt-proA)62, lacY1 or
lacZ4, glnV44, galK2 rac-(?), hisG4, rfbd1, mgl-51,
kdgK51(?), ilvC7, argE3, thi-1] (E. coli Genetic Stock
25   Center #1528) by the method of conjugation. F'16
carries the ilvGMEDAYC gene cluster. The two strains
were cross streaked on rich medium permissive for the
growth of each strain. After incubation, the plate was
replica plated to a synthetic medium containing
30   tetracycline, arginine, vitamin B1 and glucose. DE125
cannot grow on this medium because it cannot synthesize
isoleucine. Growth of AB1528 is prevented by the
inclusion of the antibiotic tetracycline and the
omission of proline and histidine from the synthetic
35   medium. A patch of cells grew on this selective medium.

70

These recombinant cells underwent single colony
isolation on the same medium. The phenotype of one
clone was determined to be Ilv$^+$, Arg$^-$, TetR, Lysine-
sensitive, male specific phage (MS2)-sensitive,
5   consistent with the simple transfer of F'16 from AB1528
to DE125. This clone was designated DE126 and has the
genotype F'16/malE52::Tn10, arg$^-$, ilvA296, thrA1101,
metL100, lysC$^+$, λ$^-$, rpsL9, malT1, xyl-7, mtl-2, thi-1?,
supE44?. It is inhibited by 20 μg/mL of L-lysine in a
10   synthetic medium.
        To select for clones from the corn cDNA library
that carried the LKR gene, 100 μL of the phagemid
library was mixed with 100 μL of an overnight culture of
DE126 grown in L broth and the cells were plated on
15   synthetic media containing vitamin B1, L-arginine,
glucose as a carbon and energy source, 100 μg/mL
ampicillin and L-lysine at 20, 30 or 40 μg/mL. Four
plates at each of the three different lysine
concentrations were prepared. The amount of phagemid
20   and DE126 cells was expected to yield about 1 x 10^5
ampicillin-resistant transfectants per plate. Ten to
thirty lysine-resistant colonies grew per plate (about 1
lysine-resistant per 5000 ampicillin-resistant
colonies).
25   Plasmid DNA was isolated from 10 independent clones
and retransformed into DE126. Seven of the ten DNAs
yielded lysine-resistant clones demonstrating that the
lysine-resistance trait was carried on the plasmid.
Several of the cloned DNAs were sequenced and
30   biochemically characterized. The inserted DNA fragments
were found to be derived from the E. coli genome, rather
than a corn cDNA indicating that the cDNA library
provided by Clontech was contaminated. A new cDNA
library will therefore be prepared and screened as
35   described above.

71

EXAMPLE 8
Construction of Synthetic Genes
in Expression Vector pSK5

To facilitate the construction and expression of
5   the synthetic genes described below, it was necessary to
construct a plasmid vector with the following
attributes:

1.   No Ear I restriction endonuclease sites
such that insertion of sequences would produce a unique
10 site.

2.   Containing a tetracycline resistance gene
to avoid loss of plasmid during growth and expression of
toxic proteins.

3.   Containing approximately 290 bp from
15 plasmid pBT430 including the T7 promoter and terminator
segment for expression of inserted sequences in E. coli.

4.   Containing unique EcoR I and Nco I
restriction endonuclease recognition sites in proper
location behind the T7 promoter to allow insertion of
20 the oligonucleotide sequences.

To obtain attributes 1 and 2 Applicants used
plasmid pSK1 which was a spontaneous mutant of pBR322
where the ampicillin gene and the Ear I site near that
gene had been deleted. Plasmid pSK1 retained the
25 tetracycline resistance gene, the unique EcoR I
restriction sites at base 1 and a single Ear I site at
base 2353. To remove the Ear I site at base 2353 of
pSK1 a polymerase chain reaction (PCR) was performed
using pSK1 as the template. Approximately 10 femtomoles
30 of pSK1 were mixed with 1 µg each of oligonucleotides
SM70 and SM71 which had been synthesized on an ABI1306B
DNA synthesizer using the manufacturer's procedures.

SM70   5'-CTGACTCGCTGCGCTCGGTC 3'   SEQ ID NO:16
35 SM71   5'-TATTTTCTCCTTACGCATCTGTGC-3' SEQ ID NO:17

72

The priming sites of these oligonucleotides on the
pSK1 template are depicted in Figure 7. The PCR was
performed using a Perkin-Elmer Cetus kit (Emeryville,
CA) according to the instructions of the vendor on a
5   thermocycler manufactured by the same company. The
25 cycles were 1 min at 95°, 2 min at 42° and 12 min at
72°. The oligonucleotides were designed to prime
replication of the entire pSK1 plasmid excluding a 30 b
fragment around the Ear I site (see Figure 7). Ten
10   microliters of the 100 μL reaction product were run on a
1% agarose gel and stained with ethidium bromide to
reveal a band of about 3.0 kb corresponding to the
predicted size of the replicated plasmid.
The remainder of the PCR reaction mix (90 μL) was
15   mixed with 20 μL of 2.5 mM deoxynucleotide triphosphates
(dATP, dTTP, dGTP, and dCTP), 30 units of Klenow enzyme
added and the mixture incubated at 37° for 30 min
followed by 65° for 10 min. The Klenow enzyme was used
to fill in ragged ends generated by the PCR. The DNA
20   was ethanol precipitated, washed with 70% ethanol, dried
under vacuum and resuspended in water. The DNA was then
treated with T4 DNA kinase in the presence of 1 mM ATP
in kinase buffer. This mixture was incubated for
30 mins at 37° followed by 10 min at 65°. To 10 μL of
25   the kinased preparation, 2 μL of 5X ligation buffer and
10 units of T4 DNA ligase were added. The ligation was
carried out at 15° for 16 h. Following ligation, the
DNA was divided in half and one half digested with Ear I
enzyme. The Klenow, kinase, ligation and restriction
30   endonuclease reactions were performed as described in
Sambrook et al., [Molecular Cloning, A Laboratory
Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory
Press]. Klenow, kinase, ligase and most restriction
endonucleases were purchased from BRL. Some restriction
35   endonucleases were purchased from NEN Biolabs (Beverly,

73

MA) or Boehringer Mannheim (Indianapolis, IN). Both the
ligated DNA samples were transformed separately into
competent JM103 [supE thi del (lac-proAB) F' [traD36
porAB, lacIq lacZ del M15] restriction minus] cells
5   using the CaCl2 method as described in Sambrook et al.,
[Molecular Cloning, A Laboratory Manual, 2nd ed. (1989)
Cold Spring Harbor Laboratory Press] and plated onto
media containing 12.5 ug/mL tetracycline. With or
without Ear I digestion the same number of transformants
10 were recovered suggesting that the Ear I site had been
removed from these constructs. Clones were screened by
preparing DNA by the alkaline lysis miniprep procedure
as described in Sambrook et al., [Molecular Cloning, A
Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor
15 Laboratory Press] followed by restriction endonuclease
digest analysis. A single clone was chosen which was
tetracycline-resistant and did not contain any Ear I
sites. This vector was designated pSK2. The remaining
EcoR I site of pSK2 was destroyed by digesting the
20 plasmid with EcoR I to completion, filling in the ends
with Klenow and ligating. A clone which did not contain
an EcoR I site was designated pSK3.

     To obtain attributes 3 and 4 above, the bacterio-
phage T7 RNA polymerase promoter/terminator segment from
25 plasmid pBT430 (see Example 2) was amplified by PCR.
Oligonucleotide primers SM78 (SEQ ID NO:18) and SM79
(SEQ ID NO:19) were designed to prime a 300b fragment
from pBT430 spanning the T7 promoter/terminator
sequences (see Figure 7).
30

SM78 5'-TTCATCGATAGGCGACCACACCCGTCC-3'   SEQ ID NO:18

SM79 5'-AATATCGATGCCACGATGCGTCCGGCG-3'   SEQ ID NO:19

35   The PCR reaction was carried out as described
previously using pBT430 as the template and a 300 bp

74

fragment was generated. The ends of the fragment were
filled in using Klenow enzyme and kinased as described
above. DNA from plasmid pSK3 was digested to completion
with PvuII enzyme and then treated with calf intestinal
5   alkaline phophatase (Boehringer Mannheim) to remove the
5' phosphate. The procedure was as described in
Sambrook et al., [Molecular Cloning, A Laboratory
Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory
Press]. The cut and phosphatased pSK3 DNA was purified
10 by ethanol precipitation and a portion used in a
ligation reaction with the PCR generated fragment
containing the T7 promoter sequence. The ligation mix
was transformed into JM103 [supE thi del (lac-proAB) F'
[traD36 porAB, lacIq lacZ del M15] restriction minus]
15 and tetracycline-resistant colonies were screened.
Plasmid DNA was prepared via the alkaline lysis mini
prep method and restriction endonuclease analysis was
performed to detect insertion and orientation of the PCR
product. Two clones were chosen for sequence analysis:
20 Plasmid pSK5 had the fragment in the orientation shown
in Figure 7. Sequence analysis performed on alkaline
denatured double-stranded DNA using Sequenase® T7 DNA
polymerase (US Biochemical Corp) and manufacturer's
suggested protocol revealed that pSK5 had no PCR
25 replication errors within the T7 promoter/terminator
sequence.

       The strategy for the construction of repeated
synthetic gene sequences based on the Ear I site is
depicted in Figure 8. The first step was the insertion
30 of an oligonucleotide sequence encoding a base gene of
14 amino acids. This oligonucleotide insert contained a
unique Ear I restriction site for subsequent insertion
of oligonucleotides encoding one or more heptad repeats
and added an unique Asp 718 restriction site for use in
35 transfer of gene sequences to plant vectors. The

75

overhanging ends of the oligonucleotide set allowed
insertion into the unique Nco I and EcoR I sites of
vector pSK5.

```
5                             M   E   E   K   M   K   A   M   E   E   K
         SM81        5'-CATGGAGGAGAAGATGAAGGCGATGGAAGAGAAG
         SM80          3'-CTCCTCTTCTACTTCCGCTACCTTCTCTTC

                     NCO I                           EAR I


10                           M   K   A              (SEQ ID NO:22)
         SM81        ATGAAGGCGTGATAGGTACCG-3'        (SEQ ID NO:20)
         SM80        TACTTCCGCACTATCCATGGCTTAA-5'    (SEQ ID NO:21)
                             ASP718 ECOR I
```

15   DNA from plasmid pSK5 was digested to completion
with Nco I and EcoR I restriction endonucleases and
purified by agarose gel electrophoresis. Purified DNA
(0.1 ug) was mixed with 1 µg of each oligonucleotide
SM80 (SEQ ID NO:14) and SM81 (SEQ ID NO:13) and ligated.
20 The ligation mixture was transformed into E. coli strain
JM103 [supE thi del (lac-proAB) F' [traD36 porAB, lacIq
lacZ del M15] restriction minus] and tetracycline
resistant transformants screened by rapid plasmid DNA
preps followed by restriction digest analysis. A clone
25 was chosen which had one each of Ear I, Nco I, Asp 718
and EcoR I sites indicating proper insertion of the
oligonucleotides. This clone was designated pSK6
(Figure 9). Sequencing of the region of DNA following
the T7 promoter confirmed insertion of oligonucleotides
30 of the expected sequence.
        Repetitive heptad coding sequences were added to
the base gene construct of described above by generating
oligonucleotide pairs which could be directly ligated
into the unique Ear I site of the base gene. Oligo-
35 nucleotides SM84 (SEQ ID NO:23) and SM85 (SEQ ID NO:24)
code for repeats of the SSP5 heptad. Oligonucleotides

76

SM82 (SEQ ID NO:25) and SM83 (SEQ ID NO:26) code for
repeats of the SSP7 heptad.

	SSP5	M E E K M K A	(SEQ ID NO:28)
5	SM84	5'-GATGGAGGAGAAGATGAAGGC-3'	(SEQ ID NO:23)
	SM85	3'- CCTCCTCTTCTACTTCCGCTA-5'	(SEQ ID NO:24)
	SSP7	M E E K L K A	(SEQ ID NO:27)
	SM82	5'-GATGGAGGAGAAGCTGAAGGC-3'	(SEQ ID NO:25)
10	SM83	3'- CCTCCTCTTCGACTTCCGCTA-5'	(SEQ ID NO:26)

        Oligonucleotide sets were ligated and purified to
obtain DNA fragments encoding multiple heptad repeats
for insertion into the expression vector. Oligonucleo-
15 tides from each set totalling about 2 µg were kinased,
and ligated for 2 h at room temperature. The ligated
multimers of the oligonucleotide sets were separated on
an 18% non-denaturing 20 X 20 X 0.015 cm polyacrylamide
gel (Acrylamide: bis-acrylamide = 19:1). Multimeric
20 forms which separated on the gel as 168 bp (8n) or
larger were purified by cutting a small piece of
polyacrylamide containing the band into fine pieces,
adding 1.0 mL of 0.5 M ammonium acetate, 1 mM EDTA
(pH 7.5) and rotating the tube at 37° overnight. The
25 polyacrylamide was spun down by centrifugation, 1 µg of
tRNA was added to the supernatant,the DNA fragments were
precipitated with 2 volumes of ethanol at -70°, washed
with 70% ethanol, dried, and resuspended in 10 µL of
water.
30   Ten micrograms of pSK6 DNA were digested to
completion with Ear I enzyme and treated with calf
intestinal alkaline phosphatase. The cut and
phosphatased vector DNA was isolated following
electrophoresis in a low melting point agarose gel by
35 cutting out the banded DNA, liquifying the agarose at
55°, and purifying over NACS PREPAC™ columns (BRL)

77

following manufacturer's suggested procedures.
Approximately 0.1 μg of purified Ear I digested and
phosphatase treated pSK6 DNA was mixed with 5 μL of the
gel purified multimeric oligonucleotide sets and
5   ligated. The ligated mixture was transformed into
E. coli strain JM103 [supE thi del (lac-proAB) F'
[traD36 porAB, lacIq lacZ del M15] restriction minus]
and tetracycline-resistant colonies selected. Clones
were screened by restriction digests of rapid plasmid
10   prep DNA to determine the length of the inserted DNA.
Restriction endonuclease analyses were usually carried
out by digesting the plasmid DNAs with Asp 718 and
Bgl II, followed by separation of fragments on 18% non-
denaturing polyacrylamide gels. Visualization of
15   fragments with ethidium bromide, showed that a 150 bp
fragment was generated when only the base gene segment
was present. Inserts of the oligonucleotide fragments
increased this size by multiples of 21 bases. From this
screening several clones were chosen for DNA sequence
20   analysis and expression of coded sequences in E. coli.
The first and last SSP5 heptads flanking the sequence of
each construct are from the base gene described above.
Inserts are designated by underlining (Table 4).

Table 4

Clone #	SEQ ID NO:	Sequence by Heptad Amino Acid Repeat (SSP)	SEQ ID NO:
C15	29	5.7.7.7.7.7.5	30
C20	31	5.7.7.7.7.5	32
C30	33	5.7.7.7.7.5	34
D16	35	5.5.5.5	36
D20	37	5.5.5.5.5	38
D33	39	5.5.5.5	40

25

Because the gel purification of the oligomeric
forms of the oligonucleotides did not give the expected

78

enrichment of longer (i.e., >8n) inserts, Applicants
used a different procedure for a subsequent round of
insertion constructions. For this series of constructs
four more sets of oligonucleotides were generated which
5   code for SSP 8,9,10 and 11 amino acid sequences
respectively:

SSP8	M E E K L K K	(SEQ ID NO:49)
SM86	5'-GATGGAGGAGAAGCTGAAGAA-3'	(SEQ ID NO:41)
SM87	3'- CCTCCTCTTCGACTTCTTCTA-5'	(SEQ ID NO:42)
SSP9	M E E K L K W	(SEQ ID NO:50)
SM88	5'-GATGGAGGAGAAGCTGAAGTG-3'	(SEQ ID NO:43)
SM89	3'- CCTCCTCTTCGACTTCACCTA-5'	(SEQ ID NO:44)
SSP10	M E E K M K K	(SEQ ID NO:51)
SM90	5'-GATGGAGGAGAAGATGAAGAA-3'	(SEQ ID NO:45)
SM91	3'- CCTCCTCTTCTACTTCTTCTA-5'	(SEQ ID NO:46)
SSP11	M E E K M K W	(SEQ ID NO:52)
SM92	5'-GATGGAGGAGAAGATGAAGTG-3'	(SEQ ID NO:47)
SM93	3'- CCTCCTCTTCTACTTCACCTA-5'	(SEQ ID NO:48)

The following HPLC procedure was used to purify
25   multimeric forms of the oligonucleotide sets after
kinasing and ligating the oligonucleotides as described
above. Chromatography was performed on a Hewlett
Packard Liquid Chromatograph instrument, Model 1090M.
Effluent absorbance was monitored at 260 nm. Ligated
30   oligonucleotides were centrifuged at 12,000xg for 5 min
and injected onto a 2.5 μ TSK DEAE-NPR ion exchange
column (35 cm x 4.6 mm I.D.) fitted with a 0.5 μ in-line
filter (Supelco). The oligonucleotides were separated
on the basis of length using a gradient elution and a
35   two buffer mobile phase [Buffer A: 25 mM Tris-Cl, pH
9.0, and Buffer B: Buffer A + 1 M NaCl]. Both Buffers
A and B were passed through 0.2 μ filters before use.

The following gradient program was used with a flow rate
of 1 mL per min at 30°:

Time	%A	%B
initial	75	25
0.5 min	55	45
5 min	50	50
20 min	38	62
23 min	0	100
30 min	0	100
31 min	75	25

5   Fractions (500 μL) were collected between 3 min and
     9 min. Fractions corresponding to lengths between
     120 bp and 2000 bp were pooled as determined from
     control separations of restriction digests of plasmid
     DNAs.

10   The 4.5 mL of pooled fractions for each oligo-
     nucleotide set were precipitated by adding 10 μg of tRNA
     and 9.0 mL of ethanol, rinsed twice with 70% ethanol and
     resuspended in 50 μL of water. Ten microliters of the
     resuspended HPLC purified oligonucleotides were added to
15   0.1 μg of the Ear I cut, phosphatased pSK6 DNA described
     above and ligated overnight at 15°. All six oligo-
     nucleotide sets described above which had been kinased
     and self-ligated but not purified by gel or HPLC were
     also used in separate ligation reactions with the pSK6
20   vector. The ligation mixtures were transformed into
     E. coli strain DH5α [supE44 del lacU169 (phi 80 lacZ del
     M15) hsdR17 recA1 endA1 gyr196 thi1 relA1] and
     tetracycline-resistant colonies selected. Applicants
     chose to use the DH5α [supE44 del lacU169 (phi 80 lacZ
25   del M15) hsdR17 recA1 endA1 gyr196 thi1 relA1] strain
     for all subsequent work because this strain has a very
     high transformation rate and is recA-. The recA-
     phenotype eliminates concerns that these repetitive DNA

structures may be substrates for homologous
recombination leading to deletion of multimeric
sequences.

5   Clones were screened as described above. Several
clones were chosen to represent insertions of each of
the six oligonucleotide sets. The first and last SSP5
heptads flanking the sequence represent the base gene
sequence. Insert sequences are underlined. Clone
numbers including the letter "H" designate HPLC-purified
10   oligonucleotides (Table 5).

Table 5

Clone #	SEQ ID NO:	Sequence by Heptad Amino Acid Repeat (SSP)	SEQ ID NO:
82-4	53	7.7.7.7.7.7.5	54
84-H3	55	5.5.5.5	56
86-H23	57	5.8.8.5	58
88-2	59	5.9.9.9.5	60
90-H8	61	5.10.10.10.5	62
92-2	63	5.11.11.5	64

The loss of the first base gene repeat in clone 82-4 may
have resulted from homologous recombination between the
15   base gene repeats 5.5 before the vector pSK6 was
transferred to the recA- strain. The HPLC procedure did
not enhance insertion of longer multimeric forms of the
oligonucleotide sets into the base gene but did serve as
an efficient purification of the ligated
20   oligonucleotides.

       Oligonucleotides were designed which coded for
mixtures of the SSP sequences and which varied codon
usage as much as possible. This was done to reduce the
possibility of deletion of repetitive inserts by
25   recombination once the synthetic genes were transformed
into plants and to extend the length of the constructed
gene segments. These oligonucleotides encode four

81

repeats of heptad coding units (28 amino acid residues)
and can be inserted at the unique Ear I site in any of
the previously constructed clones. SM96 and SM97 code
for SSP(5)₄, SM98 and SM99 code for SSP(7)₄ and SM100
plus SM101 code for SSP8.9.8.9.

```
                    M   E   E   K   M   K   A   M   E   E   K   M   K
        SM96        5'-GATGGAGGAAAAGATGAAGGCGATGGAGGAGAAAATGAAA
        SM97        3'   CCTCCTTTTCTACTTCCGCTACCTCCTCTTTTACTTT
   A   M   E   E   K   M   K   A   M   E   E   K   M   K      (SEQ ID NO:67)
  GCTATGGAGGAAAAGATGAAAGCGATGGAGGAGAAAATGAAGGC-3'            (SEQ ID NO:65)
  CGATACCTCCTTTTCTACTTTCGCTACCTCCTCTTTTACTTCCGCTA-5'        (SEQ ID NO:66)


                    M   E   E   K   L   K   A   M   E   E   K   L   K
        SM98        5'-GATGGAGGAAAAGCTGAAAGCGATGGAGGAGAAACTCAAG
        SM99        3'   CCTCCTTTTCGACTTTCGCTACCTCCTCTTTGAGTTC
   A   M   E   E   K   L   K   A   M   E   E   K   L   K      (SEQ ID NO:70)
  GCTATGGAAGAAAAGCTTAAAGCGATGGAGGAGAAACTGAAGGC-3'           (SEQ ID NO:68)
  CGATACCTTCTTTTCGAATTTCGCATCCTCCTCTTTGACTTCCGCTA-5'       (SEQ ID NO:69)


                    M   E   E   K   L   K   K   M   E   E   K   L   K
        SM100       5'-GATGGAGGAAAAGCTTAAGAAGATGGAAGAAAAGCTGAAA
        SM101       3'   CCTCCTTTTCGAATTCTTCTACCTTCTTTTCGACTTT
   W   M   E   E   K   L   K   K   M   E   E   K   L   K   W  (SEQ ID NO:73)
  TGGATGGAGGAGAAACTCAAAAAGATGGAGGAAAAGCTTAAATG-3'           (SEQ ID NO:71)
  ACCTACCTCCTCTTTGAGTTTTTCATCCTCCTTTTCGAATTTACCTA-5'       (SEQ ID NO:72)
```

DNA from clones 82-4 and 84-H3 were digested to
completion with Ear I enzyme, treated with phosphatase
and gel purified. About 0.2 µg of this DNA were mixed
with 1.0 µg of each of the oligonucleotide sets SM96 and
SM97, SM98 and SM99 or SM100 and SM101 which had been
previously kinased. The DNA and oligonucleotides were
ligated overnight and then the ligation mixes
transformed into E. coli strain DH5α. Tetracycline-
resistant colonies were screened as described above for
the presence of the oligonucleotide inserts. Clones

were chosen for sequence analysis based on their
restriction endonuclease digestion patterns (Table 6).

<u>Table 6</u>

Clone #	SEQ ID NO:	Sequence by Heptad Amino Acid Repeat (SSP)	SEQ ID NO:
2-9	74	7.7.7.7.7.7.<u>8.9.8.9</u>.5	75
3-5	78	7.7.7.7.7.7.5.5	79
5-1	76	5.5.5.<u>7.7.7.7</u>.5	77

Inserted oligonucleotide segments are underlined

5   Clone 2-9 was derived from oligonucleotides SM100
       (SEQ ID NO:71) and SM101 (SEQ ID NO:72) ligated into the
       Ear I site of clone 82-4 (see above). Clone 3-5 (SEQ ID
       NO:78) was derived from the insertion of the first 22
       bases of the oligonucleotide set SM96 (SEQ ID NO:65) and
10   SM97 (SEQ ID NO:66) into the Ear I site of clone 82-4
       (SEQ ID NO:53). This partial insertion may reflect
       improper annealing of these highly repetitive oligos.
       Clone 5-1 (SEQ ID NO:76) was derived from oligo-
       nucleotides SM98 (SEQ ID NO:68) and SM99 (SEQ ID NO:69)
15   ligated into the Ear I site of clone 84-H3 (SEQ ID
       NO:55) (see section).

       Strategy II.
            A second strategy for construction of synthetic
20   gene sequences was implemented to allow more flexibility
       in both DNA and amino acid sequence. This strategy is
       depicted in Figure 10 and Figure 11. The first step was
       the insertion of an oligonucleotide sequence encoding a
       base gene of 16 amino acids into the original vector
25   pSK5. This oligonucleotide insert contained an unique
       Ear I site as in the previous base gene construct for
       use in subsequent insertion of oligonucleotides encoding
       one or more heptad repeats. The base gene also included
       a BspH I site at the 3' terminus. The overhanging ends

83

of this cleavage site are designed to allow "in frame"
protein fusions using Nco I overhanging ends.
Therefore, gene segments can be multiplied using the
duplication scheme described in Figure 11. The
overhanging ends of the oligonucleotide set allowed
insertion into the unique Nco I and EcoR I sites of
vector pSK5.

```
                        M   E   E   K   M   K   K   L   E   E   K
SM107        5'-CATGGAGGAGAAGATGAAAAAGCTCGAAGAGAAG
SM106            3'-CTCCTCTTCTACTTTTTCGAGCTTCTCTTC
             NCO I                                EAR I


                        M   K   V   M   K              (SEQ ID NO:82)
                        ATGAAGGTCATGAAGTGATAGGTACCG-3'     (SEQ ID NO:80)
                        TACTTCCAGTACTTCACTATCCATGGCTTAA-5'  (SEQ ID NO:81)
                               BSPH I   ASP 718
```

The oligonucleotide set was inserted into pSK5 vector as
described in Strategy I above. The resultant plasmid
was designated pSK34.

Oligonucleotide sets encoding 35 amino acid
"segments" were ligated into the unique Ear I site of
the pSK34 base gene using procedures as described above.
In this case, the oligonucleotides were not gel or HPLC
purified but simply annealed and used in the ligation
reactions. The following oligonucleotide sets were
used:

```
SEG 3            L   E   E   K   M   K   A   M   E   D   K   M   K   W
SM110        5'-GCTGGAAGAAAAGATGAAGGCTATGGAGGACAAGATGAAATGG
SM111        3'-CCTTCTTTTCTACTTCCGATACCTCCTGTTCTACTTTACC


                 L   E   E   K   M   K   K              (SEQ ID NO:85)
                                                    (amino acids 8-28)
                CTTGAGGAAAAGATGAAGAA-3'              (SEQ ID NO:83)
                GAACTCCTTTTCTACTTCTTCGA-5'           (SEQ ID NO:84)
```

84

```
SEG 4          L   E   E   K   M   K   A   M   E   D   K   M   K   W
SM112      5'-GCTCGAAGAAAGATGAAGGCAATGGAAGACAAAATGAAGTGG
SM113         3'-GCTTCTTTCTACTTCCGTTACCTTCTGTTTTACTTCACC


               L   E   E   K   M   K   K                    (SEQ ID NO:86)
                                                          (amino acids 8-28)

               CTTGAGGAGAAAATGAAGAA-3'                      (SEQ ID NO:87)

               GAACTCCTCTTTTACTTCTTCGA-5'                   (SEQ ID NO:88)


SEG 5          L   K   E   E   M   A   K   M   K   D   E   M   W   K
SM114      5'-GCTCAAGGAGGAAATGGCTAAGATGAAAGACGAAATCTGGAAA
SM115         3'-GTTCCTCCTTTACCGATTCTACTTTCTGCTTTACACCTTT


               L   K   E   E   M   K   K                    (SEQ ID NO:89)
                                                          (amino acids 8-28)

               CTGAAAGAGGAAATGAAGAA                         (SEQ ID NO:90)

               GACTTTCTCCTTTACTTCTTCGA                      (SEQ ID NO:91)
```

Clones were screened for the presence of the inserted
segments by restriction digestion followed by separation
of fragments on 6% acrylamide gels. Correct insertion
of oligonucleotides was confirmed by DNA sequence
analyses. Clones containing segments 3, 4 and 5
respectively were designated pSKseg3, pSKseg4, and
pSKseg5.

These "segment" clones were used in a duplication
scheme as shown in Figure 11. Ten μg of plasmid pSKseg3
were digested to completion with Nhe I and BspH I and
the 1503 bp fragment isolated from an agarose gel using
the Whatmann paper technique. Ten μg of plasmid pSKseg4
were digested to completion with Nhe I and Nco I and the
2109 bp band gel isolated. Equal amounts of these
fragments were ligated and recombinants selected on
tetracycline. Clones were screened by restriction
digestions and their sequences confirmed. The resultant
plasmid was designated pSKseg34.

   pSKseg34 and pSKseg5 plasmid DNAs were digested,
fragments isolated and ligated in a similar manner as
above to create a plasmid containing DNA sequences
encoding segment 5 fused to segments 3 and 4. This

5 construct was designated pSKseg534 and encodes the
following amino acid sequence:

SSP534   NH2-MEEKMKKLKEEMAKMKDEMWKLKEEMKKLEEKMKVMEEKMKKLEEKMKA

          MEDKMKWLEEKMKKLEEKMKVMEEKMKKLEEKMKAMEDKMKWLEEKMKK

10   LEEKMKVMK-COOH   (SEQ ID NO:92)

                        EXAMPLE 9
        Construction of SSP Chimeric Genes for Expression
                     in the Seeds of Plants

15   To express the synthetic gene products described in
Example 8 in plant seeds, the sequences were transferred
to the seed promoter vectors CW108, CW109 or ML113
(Figure 12). The vectors CW108 and ML113 contain the
bean phaseolin promoter (from base +1 to base -494),and

20 1191 bases of the 3' sequences from bean phaseolin gene.
CW109 contains the soybean β-conglycinin promoter (from
base +1 to base -619) and the same 1191 bases of 3'
sequences from the bean phaseolin gene. These vectors
were designed to allow direct cloning of coding

25 sequences into unique Nco I and Asp 718 sites. These
vectors also provide sites (Hind III or Sal I) at the 5'
and 3' ends to allow transfer of the promoter/coding
region/ 3' sequences directly to appropriate binary
vectors.

30   To insert the synthetic storage protein gene
sequences, 10 μg of vector DNA were digested to
completion with Asp 718 and Nco I restriction
endonucleases. The linearized vector was.purified via
electrophoresis on a 1.0% agarose gel overnight

35 electrophoresis at 15 volts. The fragment was collected
by cutting the agarose in front of the band, inserting a

86

10 X 5 mm piece of Whatman 3MM paper into the agarose
and electrophoresing the fragment into the paper
[Errington, (1990) Nucleic Acids Research, 18:17].   The
fragment and buffer were spun out of the paper by
5   centrifugation and the DNA in the ~100 µL was
precipitated by adding 10 mg of tRNA, 10 µL of 3 M
sodium acetate and 200 µL of ethanol. The precipitated
DNA was washed twice with 70% ethanol and dried under
vacuum. The fragment DNA was resuspended in 20 µL of
10   water and a portion diluted 10-fold for use in ligation
reactions.

        Plasmid DNA (10 mg) from clones 3-5 and pSK534 was
digested to completion with Asp 718 and Nco I
restriction endonucleases. The digestion products were
15   separated on an 18% polyacrylamide non-denaturing gel as
described in Example 8. Gel slices containing the
desired fragments were cut from the gel and purified by
inserting the gel slices into a 1% agarose gel and
electrophoresing for 20 min at 100 volts. DNA fragments
20   were collected on 10 X 5 mm pieces of Whatman 3MM paper,
the buffer and fragments spun out by centrifugation and
the DNA precipitated with ethanol. The fragments were
resuspended in 6 µL water. One microliter of the
diluted vector fragment described above, 2 µL of 5X
25   ligation buffer and 1 µL of T4 DNA ligase were added.
The mixture was ligated overnight at 15°.

        The ligation mixes were transformed into E. coli
strain DH5a [supE44 del lacU169 (phi 80 lacZ del M15)
hsdR17 recA1 endA1 gyr196 thi1 relA1] and ampicillin-
30   resistant colonies selected. The clones were screened
by restriction endonuclease digestion analyses of rapid
plasmid DNAs and by DNA sequencing.

87

EXAMPLE 10
Tobacco Plants Containing the Chimeric Genes
Phaseolin Promoter/cts/ecodapA,
Phaseolin Promoter/cts/lysC-M4 and
β-conglycinin promoter/SSP3-5

5

The binary vector pZS97 was used to transfer the
chimeric SSP3-5 gene of Example 9 and the chimeric
E. coli dapA and lysC-M4 genes of Example 4 to tobacco
plants. Binary vector pZS97 (Figure 13) is part of a

10   binary Ti plasmid vector system [Bevan, (1984) Nucl.
Acids. Res. 12:8711-8720] of Agrobacterium tumefaciens.
The vector contains: (1) the chimeric gene nopaline
synthase::neomycin phosphotransferase (nos::NPTII) as a
selectable marker for transformed plant cells [Bevan et

15   al., (1983) Nature 304:184-186], (2) the left and right
borders of the T-DNA of the Ti plasmid [Bevan, (1984)
Nucl. Acids. Res. 12:8711-8720], (3) the E. coli lacZ
a-complementing segment [Viering et al., (1982) Gene
19:259-267] with a unique Sal I site(pSK97K) or unique

20   Hind III site (pZS97) in the polylinker region, (4) the
bacterial replication origin from the Pseudomonas
plasmid pVS1 [Itoh et al., (1984) Plasmid 11:206-220],
and (5) the bacterial β-lactamase gene as a selectable
marker for transformed A. tumefaciens.

25   Plasmid pZS97 DNA was digested to completion with
Hind III enzyme and the digested plasmid was gel
purified. The Hind III digested pZS97 DNA was mixed
with the Hind III digested and gel isolated chimeric
gene fragments, ligated, transformed as above and

30   colonies selected on ampicillin.

Binary vectors containing the chimeric genes were
transferred by tri-parental matings [Ruvkin et al.,
(1981) Nature 289:85-88] to Agrobacterium strain
LBA4404/pAL4404 [Hockema et al., (1983), Nature

35   303:179-180] selecting for carbenicillin resistance.

88

Cultures of Agrobacterium containing the binary vector
were used to transform tobacco leaf disks [Horsch et
al., (1985) Science 227:1229-1231]. Transgenic plants
were regenerated in selective medium containing
5   kanamycin.

Transformed tobacco plants containing the chimeric
gene, β-conglycinin promoter/SSP3-5/phaseolin 3' region,
were thus obtained. Two transformed lines, pSK44-3A and
pSK44-9A, which carried a single site insertion of the
10 SSP3-5 gene were identified based upon 3:1 segregation
of the marker gene for kanamycin resistance. Progeny of
the primary transformants, which were homozygous for the
transgene, pSK44-3A-6 and pSK44-9A-5, were then
identified based upon 4:0 segregation of the kanamycin
15 resistance in seeds of these plants.

Similarly, transformed tobacco plants with the
chimeric genes phaseolin 5' region/cts/lysC-M4/phaseolin
3' region and phaseolin 5' region/cts/ecodapA/phaseolin
3' region were obtained. A transformed line, BT570-45A,
20 which carried a single site insertion of the DHDPS and
AK genes was identified based upon 3:1 segregation of
the marker gene for kanamycin resistance. Progeny from
the primary transformant which were homozygous for the
transgene, BT570-45A-3 and BT570-45A-4, were then
25 identified based upon 4:0 segregation of the kanamycin
resistance in seeds of these plants.

To generate plants carrying all three chimeric
genes genetic crosses were performed using the
homozygous parents. Plants were grown to maturity in
30 greenhouse conditions. Flowers to be used as male and
female were selected one day before opening and older
flowers on the inflorescence removed. For crossing,
female flowers were chosen at the point just before
opening when the anthers were not dehiscent. The
35 corolla was opened on one side and the anthers removed.

Male flowers were chosen as flowers which had opened on
the same day and had dehiscent anthers shedding mature
pollen. The anthers were removed and used to pollinate
the pistils of the anther-stripped female flowers. The
5   pistils were then covered with plastic tubing to prevent
further pollination. The seed pods were allowed to
develop and dry for 4-6 weeks and harvested. Two to
three separate pods were recovered from each cross. The
following crosses were performed:
10

Male	X	Female
BT570-45A-3		pSK44-3A-6
BT570-45A-4		pSK44-3A-6
pSK44-3A-6		BT570-45A-4
15	BT570-45A-5		pSK44-9A-5
pSk44-9A-5		BT570-45A-5

Dried seed pods were broken open and seeds
collected and pooled from each cross. Thirty seeds were
20   counted out for each cross and for controls seeds from
selfed flowers of each parent were used. Duplicate seed
samples were hydrolyzed and assayed for total amino acid
content as described in Example 5. The amount of
increase in lysine as a percent of total seeds amino
25   acids over wild type seeds, which contain 2.56% lysine,
is presented in Table 7 along the copy number of each
gene in the endosperm of the seed.

TABLE 7

male	X	female	endosperm copy number AK & DHDPS genes	endosperm copy number SSP gene	lysine increase
BT570-45A	X	BT570-45A	1.5*	0	0
pSK44-9A	X	pSK44-9A	0	1.5*	0.12
pSK44-9A-5	X	pSK44-9A-5	0	3.0	0.29
pSK44-9A-5	X	BT570-45A-5	2	1	0.6

90

BT570-45A-5 X pSK44-9A-5	1	2	0.29
pSK44-3A X pSK44-3A	0	1.5*	0.28
pSK44-3A-6 X pSK44-3A-6	0	3.0	0.5
pSK44-3A-6 X BT570-45A-4	2	1	0.62
BT570-45A-3 X pSK44-3A-6	1	2	0.27
BT570-45A-4 X pSK44-3A-6	1	2	0.29

* copy number is average in population of seeds

The results of these crosses demonstrate that the
total lysine levels in seeds can be increased 10-25
percent by the coordinate expression of the lysine
5   biosynthesis genes and the high lysine protein SSP3-5.
In seeds derived from hybrid plants, this synergism is
strongest when the biosynthesis genes are derived from
the female parent, possibly due to gene dosage in the
endosperm. It is expected that the lysine level would
10   be further increased if the biosynthesis genes and the
lysine-rich protein genes were all homozygous.

<u>EXAMPLE 11</u>
<u>Soybean Plants Containing the Chimeric Genes</u>
<u>Phaseolin Promoter/cts/cordapA and</u>
15   <u>Phaseolin Promoter/SSP3-5</u>

Transformed soybean plants that express the
chimeric gene, phaseolin promoter/cts/cordapA/ phaseolin
3' region have been described in Example 6. Transformed
soybean plants that express the chimeric gene, phaseolin
20   promoter/SSP3-5/phaseolin 3' region, were obtained by
inserting the chimeric gene as an isolated Hind III
fragment into an equivalent soybean transformation
vector plasmid pML63 (Figure 14 Example 6) and carrying
out transformation as described in Example 6.

25   Seeds from primary transformants were sampled by
cutting small chips from the sides of the seeds away
from the embryonic axis. The chips were assayed for GUS
activity as described in Example 6 to determine which of

91

the segregating seeds carried the transgenes. Half
seeds were ground to meal and assayed for expression of
SSP3-5 protein by Enzyme Linked ImmunoSorbent Assay
(ELISA). Elisa was performed as follows:

5   A fusion protein of glutathione-S-transferase and
the SSP3-5 gene product was generated through the use of
the Pharmacia™ pGEX GST Gene Fusion System (Current
Protocols in Molecular Biology, Vol. 2, pp 16.7.1-8,
(1989) John Wiley and Sons). The fusion protein was

10   purified by affinity chromatography on glutathione
agarose (Sigma) or glutathione sepharose (Pharmacia)
beads, concentrated using Centricon 10™ (Amicon)
filters, and then subjected to SDS polyacrylamide
electrophoresis (15% Acrylamide, 19:1 Acrylamide:Bis-

15   acrylamide) for further purification. The gel was
stained with Coomassie Blue for 30 min, destained in 50%
Methanol, 10% Acetic Acid and the protein bands
electroeluted using an Amicon™ Centiluter
Microelectroeluter (Paul T. Matsudaira ed., A Practical

20   Guide to Protein and Peptide Purification for
Microsequencing, Academic Press, Inc. New York, 1989).
A second gel prepared and run in the same manner was
stained in a non acetic acid containing stain [9 parts
0.1% Coomassie Blue G250 (Bio-Rad) in 50% methanol and 1

25   part Serva Blue (Serva, Westbury, NY) in distilled
water] for 1-2 h. The gel was briefly destained in 20%
methanol, 3% glycerol for 0.5-1 h until the GST-SSP3-5
band was just barely visible. This band was excised
from the gel and sent with the electroeluted material to

30   Hazelton Laboratories for use as an antigen in
immunizing a New Zealand Rabbit. A total of 1 mg of
antigen was used (0.8 mg in gel, 0.2 mg in solution).
Test bleeds were provided by Hazelton Laboratories every
three weeks. The approximate titer was tested by

35   western blotting of E. coli extracts from cells

92

containing the SSP-3-5 gene under the control of the T7
promoter at different dilutions of protein and of serum.

IgG was isolated from the serum using a Protein A
sepharose column. The IgG was coated onto microtiter
5   plates at 5 µg per well. A separate portion of the IgG
was biotinylated.

Aqueous extracts from transgenic plants were
diluted and loaded into the wells usually starting with
a sample containing 1 µg of total protein. The sample
10   was diluted several more times to insure that at least
one of the dilutions gave a result that was within the
range of a standard curve generated on the same plate.
The standard curve was generated using chemically
synthesized SSP3-5 protein. The samples were incubated
15   for one hour at 37° and the plates washed. The
biotinylated IgG was then added to the wells. The plate
was incubated at 37° for 1 hr and washed. Alkaline
phosphatase conjugated to streptavidin was added to the
wells, incubated at 37° for 1 hr and washed. A
20   substrate consisting of 1 mg/ml p nitrophenylphosphate
in 1M diethanolamine was added to the wells and the
plates incubated at 37° for 1 hr. A 5% EDTA stop
solution was added to the wells and the absorbance read
at 405 nm minus 650 nm reading. Transgenic soybean
25   seeds contained 0.5 to 2.0% of water extractable protein
as SSP3-5.

The remaining half seeds positive for GUS and
SSP3-5 protein were planted and grown to maturity in
greenhouse conditions. To determine homozygotes for the
30   GUS phenotype, seed from these R1 plants were screened
for segregation of GUS activity as above. Plants
homozygous for the phaseolin/SSP3-5 gene were crossed
with homozygous transgenic soybeans expressing the
Corynebacterium dapA gene product.

93

As an preferred alternative to bringing the
chimeric SSP gene and chimeric cordapA geneA together
via genetic crossing a single soybean tranformation
vector carrying both genes was constructed. Plasmid
5   pML63 carrying the chimeric gene phaseolin
promoter/SSP3-5/phaseolin 3' region described above was
cleaved with restriction enzyme BamH I and the BamH I
fragment carrying the chimeric gene phaseolin
promoter/cts/cordapA/ phaseolin 3' region (Example 5)
10   was inserted. This vector can be transformed into
soybean as described in Example 6.

EXAMPLE 12

Construction of Chimeric Genes for
Expression of Corynebacterium DHDPS and SSP3-5
15   in the Embryo and Endosperm of Transformed Corn

The following chimeric genes were made for
transformation into corn:

globulin 1 promoter/mcts/cordapA/NOS 3 region
glutelin 2 promoter/mcts/cordapA/NOS 3' region
20   globulin 1 promoter/SSP3-5/globulin 1 3' region
glutelin 2 promoter/SSP3-5/10 kD 3' region

The glutelin 2 promoter was cloned from corn
genomic DNA using PCR with primers based on the
published sequence [Reina et al. (1990) Nucleic Acids
25   Res. 18:6426-6426]. The promoter fragment includes 1020
nucleotides upstream from the ATG translation start
codon. An Nco I site was introduced via PCR at the ATG
start site to allow for direct translational fusions. A
BamH I site was introduced on the 5' end of the
30   promoter. The 1.02 kb BamH I to Nco I promoter fragment
was cloned into the BamH I to Nco I sites of the plant
expression vector pML63 (see Example 11) replacing the
35S promoter to create vector pML90. This vector
contains the glutelin 2 promoter linked to the GUS
35   coding region and the NOS 3'.

The 10 kD zein 3' region was derived from a 10 kD
zein gene clone generated by PCR from genomic DNA using
oligonucleotide primers based on the published sequence
[Kirihara et al. (1988) Gene 71:359-370]. The 3' region
5   extends 940 nucleotides from the stop codon.
Restriction endonuclease sites for Kpn I, Sma I and
Xba I sites were added immediately following the TAG
stop codon by oligonucleotide insertion to facilitate
cloning. A Sma I to Hind III segment containing the
10   10 kD 3'region was isolated and ligated into Sma I and
Hind III digested pML90 to replace the NOS 3' sequence
with the 10 kD 3'region, thus creating plasmid pML103.
pML103 contains the glutelin 2 promoter, an Nco I site
at the ATG start codon of the GUS gene, Sma I and Xba I
15   sites after the stop codon, and 940 nucleotides of the
10 kD zein 3' sequence.

The globulin 1 promoter and 3' sequences were
isolated from a Clontech corn genomic DNA library using
oligonucleotide probes based on the published sequence
20   of the globulin 1 gene [Kriz et al. (1989) Plant
Physiol. 91:636]. The cloned segment includes the
promoter fragment extending 1078 nucleotides upstream
from the ATG translation start codon, the entire
globulin coding sequence including introns and the 3'
25   sequence extending 803 bases from the translational
stop. To allow replacement of the globulin 1 coding
sequence with other coding sequences an Nco I site was
introduced at the ATG start codon, and Kpn I and Xba I
sites were introduced following the translational stop
30   codon via PCR to create vector pCC50. There is a second
Nco I site within the globulin 1 promoter fragment. The
globulin 1 gene cassette is flanked by Hind III sites.

The plant amino acid biosynthetic enzymes are known
to be localized in the chloroplasts and therefore are
35   synthesized with a chloroplast targeting signal.

95

Bacterial proteins such as DHDPS have no such signal. A
chloroplast transit sequence (cts) was therefore fused
to the cordapA coding sequence in the chimeric genes
described below. For corn the cts used was based on the
5   the cts of the small subunit of ribulose 1,5-bisphos-
phate carboxylase from corn [Lebrun et al. (1987)
Nucleic Acids Res. 15:4360] and is designated acts to
distinguish it from the soybean cts. The oligo-
nucleotides SEQ ID NOS:93-98 were synthesized and used
10 essentially as described in Example 4.

     Oligonucleotides SEQ ID NO:93 and SEQ ID NO:94,
which encode the carboxy terminal part of the corn
chloroplast targeting signal, were annealed, resulting
in Xba I and Nco I compatible ends, purified via
15 polyacrylamide gel electrophoresis, and inserted into
Xba I plus Nco I digested pBT492 (see Example 3). The
insertion of the correct sequence was verified by DNA
sequencing yielding pBT556. Oligonucleotides SEQ ID
NO:95 and SEQ ID NO:96, which encode the middle part of
20 the chloroplast targeting signal, were annealed,
resulting in Bgl II and Xba I compatible ends, purified
via polyacrylamide gel electrophoresis, and inserted
into Bgl II and Xba I digested pBT556. The insertion of
the correct sequence was verified by DNA sequencing
25 yielding pBT557. Oligonucleotides SEQ ID NO:97 and SEQ
ID NO:98, which encode the amino terminal part of the
chloroplast targeting signal, were annealed, resulting
in Nco I and Afl II compatible ends, purified via
polyacrylamide gel electrophoresis, and inserted into
30 Nco I and Afl II digested pBT557. The insertion of the
correct sequence was verified by DNA sequencing yielding
pBT558. Thus the mcts was fused to the lysC-M4 gene.

     A DNA fragment containing the entire mcts was
prepared using PCR. The template DNA was pBT558 and the
35 oligonucleotide primers used were:

96

SEQ ID NO:99:
GCGCCCACCG TGATGA

5   SEQ ID NO:100:
CACCGGATTC TTCCGC

          The mcts fragment was linked to the amino terminus
     of the DHDPS protein encoded by ecodapA gene by
10   digesting with Nco I and treating with the Klenow
     fragment of DNA polymerase to fill in the 5' overhangs.
     The inserted fragment and the vector/insert junctions
     were determined to be correct by DNA sequencing,
     yielding pBT576.
15   To construct the chimeric gene:
     globulin 1 promoter/mcts/cordapA/NOS 3 region
     an Nco I to Kpn I fragment containing the mcts/ecodapA
     coding sequence was isolated from plasmid pBT576 (see
     Example 6) and inserted into Nco I plus Kpn I digested
20   pCC50 creating plasmid pBT662. Then the ecodapA coding
     sequence was replaced with the cordapA coding sequence
     as follows. An Afl II to Kpn I fragment containing the
     distal two thirds of the mcts fused to the cordapA
     coding sequence was inserted into Afl II to Kpn I
25   digested pBT662 creating plasmid pBT677.
          To construct the chimeric gene:
     glutelin 2 promoter/mcts/cordapA/NOS 3' region
     an Nco I to Kpn I fragment containing the mcts/cordapA
     coding sequence was isolated from plasmid pBT677 and
30   inserted into Nco I to Kpn I digested pML90, creating
     plasmid pBT679.
          To construct the chimeric gene:
     glutelin 2 promoter/SSP3-5/10 kD 3' region
     the plasmid pML103 (above) containing the glutelin 2
35   promoter and 10 kD zein 3' region was cleaved at the
     Nco I and Sma I sites. The SSP3-5 coding region

(Example 9) was isolated as an Nco I to blunt end
fragment by cleaving with Xba I followed by filling in
the sticky end using Klenow fragment of DNA polymerase,
then cleaving with Nco I. The 193 base pair Nco I to
5   blunt end fragment was ligated into the Nco I and Sma I
cut pML103 to create pLH104.
        To construct the chimeric gene:
glotulin 1 promoter/SSP3-5/globulin 1 3'region
the 193 base pair Nco I and Xba I fragment containing
10   the SSP3-5 coding region (Example 9) was inserted into
plasmid pCC50 (above) which had been cleaved with Xba I
to completion and then partially cut with Nco I to open
the plasmid at the ATG start codon creating pLH105.

                    EXAMPLE 13
15   Corn Plants Containing Chimeric Genes for
              Expression of Corynebacterium DHDPS
                  in the Embryo and Endosperm
        Corn was transformed with the chimeric genes:
            globulin 1 promoter/mcts/cordapA/NOS 3 region
20   or
            glutelin 2 promoter/mcts/cordapA/NOS 3' region
        Either one of two plasmid vectors containing
selectable markers were used in the transformations.
One plasmid, pDETRIC, contained the bar gene from
25   Streptomyces hygroscopicus that confers resistance to
the herbicide glufosinate [Thompson et al. (1987 The
EMBO Journal 6:2519-2523]. The bacterial gene had its
translation codon changed from GTG to ATG for proper
translation initiation in plants [De Block et al. (1987)
30   The EMBO Journal 6:2513-2518]. The bar gene was driven
by the 35S promoter from Cauliflower Mosaic Virus and
uses the termination and polyadenylation signal from the
octopine synthase gene from Agrobacterium tumefaciens.
Alternatively, the selectable marker used was 35S/Ac, a
35   synthetic phosphinothricin-N-acetyltransferase (pat)

gene under the control of the 35S promoter and 3'
terminator/polyadenylation signal from Cauliflower
Mosiac Virus [Eckes et.al., (1989) J Cell Biochem Suppl
13 D].

5   Embryogenic callus cultures were initiated from
immature embryos (about 1.0 to 1.5 mm) dissected from
kernels of a corn line bred for giving a "type II
callus" tissue culture response. The embryos were
dissected 10 to 12 d after pollination and were placed
10   with the axis-side down and in contact with agarose-
solidified N6 medium [Chu et al. (1974) Sci Sin
18:659-668] supplemented with 0.5 mg/L 2,4-D (N6-0.5).
The embryos were kept in the dark at 27°C. Friable
embryogenic callus consisting of undifferentiated masses
15   of cells with somatic proembryos and somatic embryos
borne on suspensor structures proliferated from the
scutellum of the immature embryos. Clonal embryogenic
calli isolated from individual embryos were identified
and sub-cultured on N6-0.5 medium every 2 to 3 weeks.
20   The particle bombardment method was used to
transfer genes to the callus culture cells. A
Biolistic, PDS-1000/He (BioRAD Laboratories, Hercules,
CA) was used for these experiments.
        Circular plasmid DNA or DNA which had been
25   linearized by restriction endonuclease digestion was
precipitated onto the surface of gold particles. DNA
from two or three different plasmids, one containing the
selectable marker for corn transformation, and one or
two containing the chimeric genes for increased lysine
30   accumulation in seeds were co-precipitated. To
accomplish this 1.5 µg of each DNA (in water at a
concentration of about 1 mg/mL) was added to 25 mL of
gold particles (average diameter of 1.5 µm) suspended in
water (60 mg of gold per mL). Calcium chloride (25 mL
35   of a 2.5 M solution) and spermidine (10 mL of a 1.0 M

solution) were then added to the gold-DNA suspension as the tube was vortexing. The gold particles were centrifuged in a microfuge for 10 sec and the supernatant removed. The gold particles were then

5   resuspended in 200 mL of absolute ethanol, were centrifuged again and the supernatant removed. Finally, the gold particles were resuspended in 25 mL of absolute ethanol and sonicated twice for one sec. Five μL of the DNA-coated gold particles were then loaded on each macro

10   carrier disk and the ethanol was allowed to evaporate away leaving the DNA-covered gold particles dried onto the disk.

         Embryogenic callus (from the callus line designated #LH132.5.X) was arranged in a circular area of about

15   6 cm in diameter in the center of a 100 X 20 mm petri dish containing N6-0.5 medium supplemented with 0.25M sorbitol and 0.25M mannitol. The tissue was placed on this medium for 2 h prior to bombardment as a pretreatment and remained on the medium during the

20   bombardment procedure. At the end of the 2 h pretreatment period, the petri dish containing the tissue was placed in the chamber of the PDS-1000/He. The air in the chamber was then evacuated to a vacuum of 28 inch of Hg. The macrocarrier was accelerated with a

25   helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1100 psi. The tissue was placed approximately 8 cm from the stopping screen. Four plates of tissue were bombarded with the DNA-coated gold particles. Immediately

30   following bombardment, the callus tissue was transferred to N6-0.5 medium without supplemental sorbitol or mannitol.

         Within 24 h after bombardment the tissue was transferred to selective medium, N6-0.5 medium that

35   contained 2 mg/L glufosinate and lacked casein or

proline. Tissue that continued to grow slowly on this
medium was transferred to fresh N6-0.5 medium
supplemented with glufosinate every 2 weeks. After
6-12 weeks clones of actively growing callus were
5 identified. Callus was then transferred to medium that
promotes plant regeneration.

Plants regenerated from transformed callus were
analyzed for the presence of the intact transgenes via
Southern blot or PCR. The plants were selfed or
10 outcrossed to an elite line to generate R1 or F1 seeds,
respectively. Single R1 seeds or six to eight F1 seeds
were pooled and assayed for expression of the
Corynebacterium DHDPS protein by western blot analysis.
The free amino acid composition and total amino acid
15 composition of the seeds were determined as described in
previous examples.

Expression of the Corynebacterium DHDPS protein,
driven by either the globulin or glutelin promoter, was
observed in the corn seeds (Table 8). Free lysine
20 levels in the seeds increased from about 1.4% of free
amino acids in control seeds to 15-27% in seeds
expressing Corynebacterium DHDPS from the globulin 1
promoter. The higher DHDPS expression and higher lysine
level in the selfed seed probably results from the fact
25 that half of the pooled seeds in the outcrossed lines
are expected to lack the transgene due to segregation.
A smaller increase in free lysine was observed in in
seeds expressing Corynebacterium DHDPS from the glutelin
2 promoter. Thus to increase lysine, it may be better
30 to express this enzyme in the embryo rather than the
endosperm. A high level of saccharopine, indicative of
lysine catabolism, was observed in seeds the contained
high levels of lysine.

Lysine normally represents about 2.3% of the seed
35 amino acid content. It is therefore apparent from

Table 8 that substantial increases (35%-130%) in lysine
as a percent of total seed amino acids was found in
seeds expressing Corynebacterium DHDPS from the
globulin 1 promoter.

TABLE 8

1088.1.2 line: globulin 1 promoter/mcts/cordapA/NOS 3 region
1099.2.1 line: globulin 1 promoter/mcts/cordapA/NOS 3 region
1090.2.1 line: glutelin 2 promoter/mcts/cordapA/NOS 3' region

TRANSGENIC LINE	WESTERN CORYNE. DHDPS	% LYS of FREE SEED AMINO ACIDS	% LYS of TOTAL SEED AMINO ACIDS
1088.1.2 x elite	+	15	3.1
1099.2.1 selfed	++	27	5.3
1090.2.1 x elite	+	2.3	1.7

EXAMPLE 14

Transformation of Soybean with the Kunitz trypsin inhibitor 3 promoter/cts/cordapA Chimeric Gene

A seed-specific expression cassette composed of the
promoter and transcription terminator from the the
soybean Kunitz trypsin inhibitor 3 (KTI3) gene [Jofuku
et al. (1989) Plant Cell 1:427-435] was created. The
KTI3 cassette includes about 2000 nucleotides upstream
(5') from the translation initiation codon and about 200
nucleotides downstream (3') from the translation stop
codon of Kunitz trypsin inhibitor 3. Between the 5' and
3' regions restriction endonuclease sites Nco I (which
includes the ATG translation initiation codon) and Kpn I
were created to permit insertion of the Corynebacterium
dapA gene. The entire cassette was flanked by BamH I
and Sal I sites.

As described in Example 4 a chloroplast transit
sequence (cts) was fused to the dapA coding sequence in
the chimeric gene. The cts used was based on the the
cts of the small subunit of ribulose 1,5-bisphosphate
carboxylase from soybean [Berry-Lowe et al. (1982) J.

102

Mol. Appl. Genet. 1:483-498]. A 1030 bp Nco I-Kpn I
fragment containing the cts attached to the cordapA
coding region was isolated from an agarose gel following
electrophoresis and inserted into the KTI3 expression
5   cassette yielding plasmid pML102 (Figure 15).

Plasmid pML102 was introduced into soybean by
particle-mediated bombardment by Agracetus Company
(Middleton, WI), according to the procedure described in
United States Patent No. 5,015,580. To screen for
10   transformed cells, plasmid pML102 was co-bombarded with
another plasmid carrying a soybean transformation marker
gene consisting of the 35S promoter from Cauliflower
Mosaic Virus driving expression of the E. coli
β-glucuronidase (GUS) gene [Jefferson et al. (1986)
15   Proc. Natl. Acad. Sci. USA 83:8447-8451] with the Nos 3'
region.

It was expected that the transgenes would be
segregating in the R1 seeds of the transformed plants.
To identify seeds that carried the transformation marker
20   gene, a small chip of the seed was cut off with a razor
and put into a well in a disposable plastic microtiter
plate. A GUS assay mix consisting of 100 mM NaH_2PO_4,
10 mM EDTA, 0.5 mM $K_4Fe(CN)_6$, 0.1% Triton X-100,
0.5 mg/mL 5-Bromo-4-chloro-3-indolyl β-D-glucuronic acid
25   was prepared and 0.15 mL was added to each microtiter
well. The microtiter plate was incubated at 37° for
45 minutes. The development of blue color indicated the
expression of GUS in the seed.

To measure the total amino acid composition of
30   mature seeds, 1-1.4 milligrams of the seed meal was
hydrolyzed in 6N hydrochloric acid, 0.4% β-mercapto-
ethanol under nitrogen for 24 h at 110-120°C; 1/50 of
the sample was run on a Beckman Model 6300 amino acid
analyzer using post-column ninhydrin detection. Lysine
35   (and other amino acid) levels in the seeds were compared

as percentages of the total amino acids. Wild type
soybean seeds contain 5.7-6.0% lysine.

     One hundred fifty individual seeds from sixteen
independent transformed lines were analyzed (Table 9).
5 Ten of the sixteen lines had seeds with a lysine content
of 7% of the total seed amino acids or greater, a 16-22%
increase over wild type seeds. Thus, more than 62% of
the transformation events had co-integrated the plasmid
carrying the cordapA gene along with the plasmid bearing
10 the marker GUS gene. About 80% of the high lysine seeds
were GUS positive, suggesting that the plasmid carrying
the cordapA gene usually integrated at the same
chromosomal site as the plasmid carrying the GUS gene.
However, in some transformed lines, e.g. 260-05, there
15 was little correlation between the GUS positive and high
lysine phenotypes, indicating that the two plasmids
integrated at unlinked sites. Both of these types of
transformation events were expected based upon the
procedure used for this transformation.
20   Seeds with a lysine content greater than 20% of the
total seed amino acids were obtained. This represents
nearly a three hundred percent increase in seed lysine
content.

TABLE 9

LINE #	SEED #	GUS	%LYS
257-1	G1	+	8.30
	G2	+	7.99
	G3	+	11.51
	G4	+	8.52
	G33	+	7.68
	G34	+	9.93
	G35	−	5.97
	G36	−	5.71
	G37	+	7.48

104

	G38	+	9.42
	G39	+	10.44
	G40	+	8.63
	G41	+	9.42
	G42	+	8.53
	G43	+	10.54
	G44	−	5.83
	G45	+	7.15
	G46	+	7.85
	G47	+	7.34
257−21	G21	+	12.90
	G22	+	11.52
	G23	+	9.34
	G24	−	5.82
	G25	−	5.61
	G26	−	5.70
	G27	−	5.84
	G28	−	14.27
	G48	+	15.23
	G49	+	18.79
	G50	+	13.82
	G51	−	5.94
	G52	+	13.29
	G53	+	14.61
257−41	G54	+	6.28
	G55	+	6.27
	G56	+	6.32
	G57	+	6.4
	G180	+	5.75
	G181	+	7.42
257−46	G60	+	6.76
	G61	+	6.73
	G62	−	6.18

	G63	+	6.13
	G182	+	6.83
	G183	+	6.23
257–49	G78	−	6.40
	G79	+	6.46
	G184	+	6.37
	G185	+	6.15
	G186	+	6.41
	G187	+	7.90
257–50	G88	−	6.15
	G89	−	6.12
	G188	+	6.19
	G189	+	6.07
	G190	+	6.09
	G191	+	6.30
257–51	G228	−	5.81
	G229	−	5.74
	G230	−	5.59
	G231	−	6.00
	G232	−	5.89
	G233	+	21.49
	G234	+	20.30
	G235	+	11.89
	G236	+	12.40
	G237	+	15.09
	G238	+	12.79
	G239	+	17.19
260–05	G90	−	5.41
	G91	−	7.65
	G95	−	6.39
	G96	−	5.80
	G97	−	6.12

106

	G98	–	5.90
	G99	–	6.17
	G160	–	8.04
	G161	–	12.64
	G162	–	6.91
	G163	–	5.83
	G164	–	8.28
	G165	–	12.52
	G166	–	5.68
	G167	–	9.92
	G168	–	5.89
	G169	–	6.10
	G170	+	6.49
	G171	+	6.10
	G172	–	12.83
	G173	–	6.55
	G174	–	6.62
	G175	+	13.02
	G176	–	10.13
	G177	–	5.97
	G178	–	11.37
	G179	–	12.63
260-13	G108	+	6.64
	G109	+	7.92
	G192	+	10.29
	G193	+	7.37
	G194	+	6.73
	G195	+	10.35
260-16	G29	+	11.64
	G30	+	14.87
	G31	+	15.02
	G32	–	6.24

260-17	G115	+	11.91
	G116	−	6.21
	G117	−	6.08
	G118	−	6.28
	G119	−	6.30
	G196	+	7.76
260-23	G129	+	5.93
	G197	+	6.04
	G198	+	5.99
	G199	+	6.11
	G200	+	6.35
	G201	+	6.19
260-31	G202	+	6.19
	G203	+	6.19
	G204	+	6.13
	G205	+	6.40
	G206	+	6.73
	G207	+	6.23
260-33	G217	+	6.80
	G218	−	7.00
	G219	−	6.80
	G220	−	6.10
	G221	−	6.83
	G222	−	6.18
	G223	+	5.92
	G224	+	6.61
	G226	+	6.17
	G227	+	6.43
	G240	+	6.25
	G241	+	6.13
260-44	G148	+	6.51
	G149	+	6.21

108

	G208	+	6.02
	G209	+	6.17
	G210	+	6.12
	G211	+	6.09
260-46	G158	-	6.00
	G159	+	6.30
	G212	+	6.40
	G213	+	6.50
	G214	+	6.40
	G215	+	6.60

109

<u>SEQUENCE LISTING</u>

(1)   GENERAL INFORMATION:

    (i)   APPLICANT:
        (A) NAME: E. I. DU PONT DE NEMOURS AND
                COMPANY
        (B) STREET: 1007 MARKET STREET
        (C) CITY: WILMINGTON
        (D) STATE: DELAWARE
        (E) COUNTRY: UNITED STATES OF AMERICA
        (F) POSTAL CODE (ZIP): 19898
        (G) TELEPHONE: 302-992-4931
        (H) TELEFAX: 302-773-0164
        (I) TELEX: 6717325

    (ii)   TITLE OF INVENTION: CHIMERIC GENES AND
                         METHODS FOR INCREASING
                         THE LYSINE CONTENT OF
                         THE SEEDS OF CORN,
                         SOYBEAN AND RAPESEED
                         PLANTS

    (iii) NUMBER OF SEQUENCES: 100

    (iv)   COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: DISKETTE, 3.50 INCH
        (B) COMPUTER: MACINTOSH
        (C) OPERATING SYSTEM: MACINTOSH, 6.0
        (D) SOFTWARE: MICROSOFT WORD, 4.0

    (v)   CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER:
        (B) FILING DATE:
        (C) CLASSIFICATION:

    (vi)   PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: 08/160,117
        (B) FILING DATE: NOVEMBER 30, 1993

    (vii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: BARBARA C. SIEGELL
        (B) REGISTRATION NUMBER: 30,684
        (C) REFERENCE/DOCKET NUMBER: BB-1055-B

110

```
(2)   INFORMATION FOR SEQ ID NO:1:

      (i)   SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  48 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

     (ii)   MOLECULE TYPE:  DNA (genomic)

     (xi)   SEQUENCE DESCRIPTION:  SEQ ID NO:1:

CCCGGGCCAT GGCTACAGGT TTAACAGCTA AGACCGGAGT AGAGCACT   48

(2)   INFORMATION FOR SEQ ID NO:2:

      (i)   SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  37 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

     (ii)   MOLECULE TYPE:  DNA (genomic)

     (xi)   SEQUENCE DESCRIPTION:  SEQ ID NO:2:

GATATCGAAT TCTCATTATA GAACTCCAGC TTTTTTC                  37

(2)   INFORMATION FOR SEQ ID NO:3:

      (i)   SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  917 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

     (ii)   MOLECULE TYPE:  DNA (genomic)

     (ix)   FEATURE:
            (A)   NAME/KEY:  CDS
            (B)   LOCATION:  3..911

     (xi)   SEQUENCE DESCRIPTION:  SEQ ID NO:3:

CC ATG GCT ACA GGT TTA ACA GCT AAG ACC GGA GTA GAG CAC TTC GGC   47
   Met Ala Thr Gly Leu Thr Ala Lys Thr Gly Val Glu His Phe Gly
   1               5                  10                  15

ACC GTT GGA GTA GCA ATG GTT ACT CCA TTC ACG GAA TCC GGA GAC ATC   95
Thr Val Gly Val Ala Met Val Thr Pro Phe Thr Glu Ser Gly Asp Ile
            20                  25                  30
```

111

```
GAT ATC GCT GCT GGC CGC GAA GTC GCG GCT TAT TTG GTT GAT AAG GGC   143
Asp Ile Ala Ala Gly Arg Glu Val Ala Ala Tyr Leu Val Asp Lys Gly
            35                  40                  45

TTG GAT TCT TTG GTT CTC GCG GGC ACC ACT GGT GAA TCC CCA ACG ACA   191
Leu Asp Ser Leu Val Leu Ala Gly Thr Thr Gly Glu Ser Pro Thr Thr
            50                  55                  60

ACC GCC GCT GAA AAA CTA GAA CTG CTC AAG GCC GTT CGT GAG GAA GTT   239
Thr Ala Ala Glu Lys Leu Glu Leu Leu Lys Ala Val Arg Glu Glu Val
    65                  70                  75

GGG GAT CGG GCG AAG CTC ATC GCC GGT GTC GGA ACC AAC AAC ACG CGG   287
Gly Asp Arg Ala Lys Leu Ile Ala Gly Val Gly Thr Asn Asn Thr Arg
80                  85                  90                  95

ACA TCT GTG GAA CTT GCG GAA GCT GCT GCT TCT GCT GGC GCA GAC GGC   335
Thr Ser Val Glu Leu Ala Glu Ala Ala Ala Ser Ala Gly Ala Asp Gly
                100                 105                 110

CTT TTA GTT GTA ACT CCT TAT TAC TCC AAG CCG AGC CAA GAG GGA TTG   383
Leu Leu Val Val Thr Pro Tyr Tyr Ser Lys Pro Ser Gln Glu Gly Leu
                115                 120                 125

CTG GCG CAC TTC GGT GCA ATT GCT GCA GCA ACA GAG GTT CCA ATT TGT   431
Leu Ala His Phe Gly Ala Ile Ala Ala Ala Thr Glu Val Pro Ile Cys
            130                 135                 140

CTC TAT GAC ATT CCT GGT CGG TCA GGT ATT CCA ATT GAG TCT GAT ACC   479
Leu Tyr Asp Ile Pro Gly Arg Ser Gly Ile Pro Ile Glu Ser Asp Thr
            145                 150                 155

ATG AGA CGC CTG AGT GAA TTA CCT ACG ATT TTG GCG GTC AAG GAC GCC   527
Met Arg Arg Leu Ser Glu Leu Pro Thr Ile Leu Ala Val Lys Asp Ala
160                 165                 170                 175

AAG GGT GAC CTC GTT GCA GCC ACG TCA TTG ATC AAA GAA ACG GGA CTT   575
Lys Gly Asp Leu Val Ala Ala Thr Ser Leu Ile Lys Glu Thr Gly Leu
                180                 185                 190

GCC TGG TAT TCA GGC GAT GAC CCA CTA AAC CTT GTT TGG CTT GCT TTG   623
Ala Trp Tyr Ser Gly Asp Asp Pro Leu Asn Leu Val Trp Leu Ala Leu
            195                 200                 205

GGC GGA TCA GGT TTC ATT TCC GTA ATT GGA CAT GCA GCC CCC ACA GCA   671
Gly Gly Ser Gly Phe Ile Ser Val Ile Gly His Ala Ala Pro Thr Ala
            210                 215                 220

TTA CGT GAG TTG TAC ACA AGC TTC GAG GAA GGC GAC CTC GTC CGT GCG   719
Leu Arg Glu Leu Tyr Thr Ser Phe Glu Glu Gly Asp Leu Val Arg Ala
            225                 230                 235

CGG GAA ATC AAC GCC AAA CTA TCA CCG CTG GTA GCT GCC CAA GGT CGC   767
Arg Glu Ile Asn Ala Lys Leu Ser Pro Leu Val Ala Ala Gln Gly Arg
240                 245                 250                 255
```

112

```
TTG GGT GGA GTC AGC TTG GCA AAA GCT GCT CTG CGT CTG CAG GGC ATC   815
Leu Gly Gly Val Ser Leu Ala Lys Ala Ala Leu Arg Leu Gln Gly Ile
            260             265             270

AAC GTA GGA GAT CCT CGA CTT CCA ATT ATG GCT CCA AAT GAG CAG GAA   863
Asn Val Gly Asp Pro Arg Leu Pro Ile Met Ala Pro Asn Glu Gln Glu
            275             280             285

CTT GAG GCT CTC CGA GAA GAC ATG AAA AAA GCT GGA GTT CTA TAA TGAGAATTC   918
Leu Glu Ala Leu Arg Glu Asp Met Lys Lys Ala Gly Val Leu  *
            290             295             300
```

    (2)   INFORMATION FOR SEQ ID NO:4:

        (i) SEQUENCE CHARACTERISTICS:
           (A)   LENGTH: 22 base pairs
           (B)   TYPE: nucleic acid
           (C)   STRANDEDNESS: single
           (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CTTCCCGTGA CCATGGGCCA TC   22

    (2)   INFORMATION FOR SEQ ID NO:5:

        (i) SEQUENCE CHARACTERISTICS:
           (A)   LENGTH: 1350 base pairs
           (B)   TYPE: nucleic acid
           (C)   STRANDEDNESS: single
           (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (ix) FEATURE:
           (A)   NAME/KEY: CDS
           (B)   LOCATION: 1..1350

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATG GCT GAA ATT GTT GTC TCC AAA TTT GGC GGT ACC AGC GTA GCT GAT   48
Met Ala Glu Ile Val Val Ser Lys Phe Gly Gly Thr Ser Val Ala Asp
1               5               10              15

TTT GAC GCC ATG AAC CGC AGC GCT GAT ATT GTG CTT TCT GAT GCC AAC   96
Phe Asp Ala Met Asn Arg Ser Ala Asp Ile Val Leu Ser Asp Ala Asn
            20              25              30

GTG CGT TTA GTT GTC CTC TCG GCT TCT GCT GGT ATC ACT AAT CTG CTG   144
Val Arg Leu Val Val Leu Ser Ala Ser Ala Gly Ile Thr Asn Leu Leu
            35              40              45
```

```
GTC GCT TTA GCT GAA GGA CTG GAA CCT GGC GAG CGA TTC GAA AAA CTC    192
Val Ala Leu Ala Glu Gly Leu Glu Pro Gly Glu Arg Phe Glu Lys Leu
        50              55              60

GAC GCT ATC CGC AAC ATC CAG TTT GCC ATT CTG GAA CGT CTG CGT TAC    240
Asp Ala Ile Arg Asn Ile Gln Phe Ala Ile Leu Glu Arg Leu Arg Tyr
 65              70              75              80

CCG AAC GTT ATC CGT GAA GAG ATT GAA CGT CTG CTG GAG AAC ATT ACT    288
Pro Asn Val Ile Arg Glu Glu Ile Glu Arg Leu Leu Glu Asn Ile Thr
                85              90              95

GTT CTG GCA GAA GCG GCG GCG CTG GCA ACG TCT CCG GCG CTG ACA GAT    336
Val Leu Ala Glu Ala Ala Ala Leu Ala Thr Ser Pro Ala Leu Thr Asp
            100             105             110

GAG CTG GTC AGC CAC GGC GAG CTG ATG TCG ACC CTG CTG TTT GTT GAG    384
Glu Leu Val Ser His Gly Glu Leu Met Ser Thr Leu Leu Phe Val Glu
            115             120             125

ATC CTG CGC GAA CGC GAT GTT CAG GCA CAG TGG TTT GAT GTA CGT AAA    432
Ile Leu Arg Glu Arg Asp Val Gln Ala Gln Trp Phe Asp Val Arg Lys
        130             135             140

GTG ATG CGT ACC AAC GAC CGA TTT GGT CGT GCA GAG CCA GAT ATA GCC    480
Val Met Arg Thr Asn Asp Arg Phe Gly Arg Ala Glu Pro Asp Ile Ala
145             150             155             160

GCG CTG GCG GAA CTG GCC GCG CTG CAG CTG CTC CCA CGT CTC AAT GAA    528
Ala Leu Ala Glu Leu Ala Ala Leu Gln Leu Leu Pro Arg Leu Asn Glu
            165             170             175

GGC TTA GTG ATC ACC CAG GGA TTT ATC GGT AGC GAA AAT AAA GGT CGT    576
Gly Leu Val Ile Thr Gln Gly Phe Ile Gly Ser Glu Asn Lys Gly Arg
            180             185             190

ACA ACG ACG CTT GGC CGT GGA GGC AGC GAT TAT ACG GCA GCC TTG CTG    624
Thr Thr Thr Leu Gly Arg Gly Gly Ser Asp Tyr Thr Ala Ala Leu Leu
            195             200             205

GCG GAG GCT TTA CAC GCA TCT CGT GTT GAT ATC TGG ACC GAC GTC CCG    672
Ala Glu Ala Leu His Ala Ser Arg Val Asp Ile Trp Thr Asp Val Pro
210             215             220

GGC ATC TAC ACC ACC GAT CCA CGC GTA GTT TCC GCA GCA AAA CGC ATT    720
Gly Ile Tyr Thr Thr Asp Pro Arg Val Val Ser Ala Ala Lys Arg Ile
225             230             235             240

GAT GAA ATC GCG TTT GCC GAA GCG GCA GAG ATG GCA ACT TTT GGT GCA    768
Asp Glu Ile Ala Phe Ala Glu Ala Ala Glu Met Ala Thr Phe Gly Ala
            245             250             255

AAA GTA CTG CAT CCG GCA ACG TTG CTA CCC GCA GTA CGC AGC GAT ATC    816
Lys Val Leu His Pro Ala Thr Leu Leu Pro Ala Val Arg Ser Asp Ile
            260             265             270
```

114

```
CCG GTC TTT GTC GGC TCC AGC AAA GAC CCA CGC GCA GGT GGT ACG CTG    864
Pro Val Phe Val Gly Ser Ser Lys Asp Pro Arg Ala Gly Gly Thr Leu
        275                 280                 285

GTG TGC AAT AAA ACT GAA AAT CCG CCG CTG TTC CGC GCT CTG GCG CTT    912
Val Cys Asn Lys Thr Glu Asn Pro Pro Leu Phe Arg Ala Leu Ala Leu
    290                 295                 300

CGT CGC AAT CAG ACT CTG CTC ACT TTG CAC AGC CTG AAT ATG CTG CAT    960
Arg Arg Asn Gln Thr Leu Leu Thr Leu His Ser Leu Asn Met Leu His
305                 310                 315                 320

TCT CGC GGT TTC CTC GCG GAA GTT TTC GGC ATC CTC GCG CGG CAT AAT   1008
Ser Arg Gly Phe Leu Ala Glu Val Phe Gly Ile Leu Ala Arg His Asn
                325                 330                 335

ATT TCG GTA GAC TTA ATC ACC ACG TCA GAA GTG AGC GTG GCA TTA ACC   1056
Ile Ser Val Asp Leu Ile Thr Thr Ser Glu Val Ser Val Ala Leu Thr
            340                 345                 350

CTT GAT ACC ACC GGT TCA ACC TCC ACT GGC GAT ACG TTG CTG ACG CAA   1104
Leu Asp Thr Thr Gly Ser Thr Ser Thr Gly Asp Thr Leu Leu Thr Gln
            355                 360                 365

TCT CTG CTG ATG GAG CTT TCC GCA CTG TGT CGG GTG GAG GTG GAA GAA   1152
Ser Leu Leu Met Glu Leu Ser Ala Leu Cys Arg Val Glu Val Glu Glu
    370                 375                 380

GGT CTG GCG CTG GTC GCG TTG ATT GGC AAT GAC CTG TCA AAA GCC TGC   1200
Gly Leu Ala Leu Val Ala Leu Ile Gly Asn Asp Leu Ser Lys Ala Cys
385                 390                 395                 400

GCC GTT GGC AAA GAG GTA TTC GGC GTA CTG GAA CCG TTC AAC ATT CGC   1248
Ala Val Gly Lys Glu Val Phe Gly Val Leu Glu Pro Phe Asn Ile Arg
                405                 410                 415

ATG ATT TGT TAT GGC GCA TCC AGC CAT AAC CTG TGC TTC CTG GTG CCC   1296
Met Ile Cys Tyr Gly Ala Ser Ser His Asn Leu Cys Phe Leu Val Pro
                420                 425                 430

GGC GAA GAT GCC GAG CAG GTG GTG CAA AAA CTG CAT AGT AAT TTG TTT   1344
Gly Glu Asp Ala Glu Gln Val Val Gln Lys Leu His Ser Asn Leu Phe
            435                 440                 445

GAG TAA                                                            1350
Glu *
    450
```

        (2)   INFORMATION FOR SEQ ID NO:6:

            (i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 36 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

            (ii)   MOLECULE TYPE: DNA (genomic)

115

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCCATGGC TGAAATTGTT GTCTCCAAAT TTGGCG   36

(2)   INFORMATION FOR SEQ ID NO:7:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 36 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTACCGCCAA ATTTGGAGAC AACAATTTCA GCCATG   36

(2)   INFORMATION FOR SEQ ID NO:8:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 75 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CATGGCTGGC TTCCCCACGA GGAAGACCAA CAATGACATT ACCTCCATTG CTAGCAACGG   60

TGGAAGAGTA CAATG   75

(2)   INFORMATION FOR SEQ ID NO:9:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 75 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CATGCATTGT ACTCTTCCAC CGTTGCTAGC AATGGAGGTA ATGTCATTGT TGGTCTTCCT   60

CGTGGGGAAG CCAGC   75

    (2)   INFORMATION FOR SEQ ID NO:10:

           (i) SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 90 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA (genomic)

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CATGGCTTCC TCAATGATCT CCTCCCCAGC TGTTACCACC GTCAACCGTG CCGGTGCCGG   60

CATGGTTGCT CCATTCACCG GCCTCAAAAG   90

    (2)   INFORMATION FOR SEQ ID NO:11:

           (i) SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 90 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA (genomic)

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CATGCTTTTG AGGCCGGTGA ATGGAGCAAC CATGCCGGCA CCGGCACGGT TGACGGTGGT   60

AACAGCTGGG GAGGAGATCA TTGAGGAAGC   90

    (2)   INFORMATION FOR SEQ ID NO:12:

           (i) SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 23 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA (genomic)

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCGGTTTGCT GTAATAGGTA CCA   23

    (2)   INFORMATION FOR SEQ ID NO:13:

           (i) SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 31 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

117

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGCTTGGTAC CTATTACAGC AAACCGGCAT G   31

(2)   INFORMATION FOR SEQ ID NO:14:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 27 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCTTCCTCAA TGATCTCCTC CCCAGCT   27

(2)   INFORMATION FOR SEQ ID NO:15:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 28 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CATTGTACTC TTCCACCGTT GCTAGCAA   28

(2)   INFORMATION FOR SEQ ID NO:16:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 20 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
            (A)   NAME/KEY: misc_feature
            (B)   LOCATION: 1..20
            (D)   OTHER INFORMATION: /product= "synthetic
                                       oligonucleotide"
                                       /standard_name= "SM
                                       70"

118

(xi)   SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTGACTCGCT GCGCTCGGTC   20

(2)   INFORMATION FOR SEQ ID NO:17:

        (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH: 24 base pairs
              (B)   TYPE: nucleic acid
              (C)   STRANDEDNESS: single
              (D)   TOPOLOGY: linear

       (ii)   MOLECULE TYPE: DNA (genomic)

       (ix)   FEATURE:
              (A)   NAME/KEY: misc_feature
              (B)   LOCATION: 1..24
              (D)   OTHER INFORMATION: /product= "synthetic
                                         oligonucleotide"
                                         /standard_name= "SM
                                         71"

       (xi)   SEQUENCE DESCRIPTION: SEQ ID NO:17:

TATTTTCTCC TTACGCATCT GTGC   24

(2)   INFORMATION FOR SEQ ID NO:18:

        (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH: 27 base pairs
              (B)   TYPE: nucleic acid
              (C)   STRANDEDNESS: single
              (D)   TOPOLOGY: linear

       (ii)   MOLECULE TYPE: DNA (genomic)

       (ix)   FEATURE:
              (A)   NAME/KEY: misc_feature
              (B)   LOCATION: 1..27
              (D)   OTHER INFORMATION: /product= "synthetic
                                         oligonucleotide"
                                         /standard_name= "SM
                                         78"

       (xi)   SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTCATCGATA GGCGACCACA CCCGTCC   27

(2)   INFORMATION FOR SEQ ID NO:19:

        (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH: 27 base pairs
              (B)   TYPE: nucleic acid

119

```
(C)    STRANDEDNESS:  single
(D)    TOPOLOGY:  linear

(ii)  MOLECULE TYPE:  DNA (genomic)

(ix)  FEATURE:
(A)    NAME/KEY:  misc_feature
(B)    LOCATION:  1..27
(D)    OTHER INFORMATION:  /product= "synthetic
                           oligonucleotide"
                           /standard_name= "SM
                           79"

(xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:19:
```

AATATCGATG CCACGATGCG TCCGGCG   27

```
(2)    INFORMATION FOR SEQ ID NO:20:

(i)  SEQUENCE CHARACTERISTICS:
(A)    LENGTH:  55 base pairs
(B)    TYPE:  nucleic acid
(C)    STRANDEDNESS:  single
(D)    TOPOLOGY:  linear

(ii)  MOLECULE TYPE:  DNA (genomic)

(ix)  FEATURE:
(A)    NAME/KEY:  misc_feature
(B)    LOCATION:  1..55
(D)    OTHER INFORMATION:  /product= "synthetic
                           oligonucleotide"
                           /standard_name= "SM
                           81"

(xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:20:
```

CATGGAGGAG AAGATGAAGG CGATGGAAGA GAAGATGAAG GCGTGATAGG TACCG   55

```
(2)    INFORMATION FOR SEQ ID NO:21:

(i)  SEQUENCE CHARACTERISTICS:
(A)    LENGTH:  55 base pairs
(B)    TYPE:  nucleic acid
(C)    STRANDEDNESS:  single
(D)    TOPOLOGY:  linear

(ii)  MOLECULE TYPE:  DNA (genomic)

(ix)  FEATURE:
(A)    NAME/KEY:  misc_feature
(B)    LOCATION:  1..55
```

            (D)   OTHER INFORMATION: /product= "synthetic
                                         oligonucleotide"
                                         /standard_name= "SM
                                         80"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AATTCGGTAC CTATCACGCC TTCATCTTCT CTTCCATCGC CTTCATCTTC TCCTC   55

    (2)   INFORMATION FOR SEQ ID NO:22:

           (i)   SEQUENCE CHARACTERISTICS:
                  (A)   LENGTH: 14 amino acids
                  (B)   TYPE: amino acid
                  (C)   STRANDEDNESS: unknown
                  (D)   TOPOLOGY: unknown

          (ii) MOLECULE TYPE: protein

          (ix) FEATURE:
                  (A)   NAME/KEY: Protein
                  (B)   LOCATION: 1..14
                  (D)   OTHER INFORMATION: /label= name
                                            /note= "base gene
                                            [(SSP5)2]"

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
1   5   10

    (2)   INFORMATION FOR SEQ ID NO:23:

           (i)   SEQUENCE CHARACTERISTICS:
                  (A)   LENGTH: 21 base pairs
                  (B)   TYPE: nucleic acid
                  (C)   STRANDEDNESS: single
                  (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA (genomic)

          (ix) FEATURE:
                  (A)   NAME/KEY: misc_feature
                  (B)   LOCATION: 1..21
                  (D)   OTHER INFORMATION: /product= "synthetic
                                            oligonucleotide"
                                            /standard_name= "SM
                                            84"

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GATGGAGGAG AAGATGAAGG C   21

121

(2)   INFORMATION FOR SEQ ID NO:24:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 21 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
            (A)   NAME/KEY: misc_feature
            (B)   LOCATION: 1..21
            (D)   OTHER INFORMATION: /product= "synthetic
                                     oligonucleotide"
                                     /standard_name= "SM
                                     85"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATCGCCTTCA TCTTCTCCTC C   21

(2)   INFORMATION FOR SEQ ID NO:25:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 21 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
            (A)   NAME/KEY: misc_feature
            (B)   LOCATION: 1..21
            (D)   OTHER INFORMATION: /product= "synthetic
                                     oligonucleotide"
                                     /standard_name= "SM
                                     82"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GATGGAGGAG AAGCTGAAGG C   21

(2)   INFORMATION FOR SEQ ID NO:26:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 21 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

122

```
    (ix) FEATURE:
         (A)  NAME/KEY: misc_feature
         (B)  LOCATION: 1..21
         (D)  OTHER INFORMATION: /product= "synthetic
                                 oligonucleotide"
                                 /standard_name= "SM
                                 83"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATCGCCTTCA GCTTCTCCTC C                                    21

(2)  INFORMATION FOR SEQ ID NO:27:

    (i) SEQUENCE CHARACTERISTICS:
         (A)  LENGTH:  7 amino acids
         (B)  TYPE:  amino acid
         (C)  STRANDEDNESS:  unknown
         (D)  TOPOLOGY:  unknown

    (ii) MOLECULE TYPE:  protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met Glu Glu Lys Leu Lys Ala
1               5

(2)  INFORMATION FOR SEQ ID NO:28:

    (i) SEQUENCE CHARACTERISTICS:
         (A)  LENGTH:  7 amino acids
         (B)  TYPE:  amino acid
         (C)  STRANDEDNESS:  unknown
         (D)  TOPOLOGY:  unknown

    (ii) MOLECULE TYPE:  protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Glu Glu Lys Met Lys Ala
1               5

(2)  INFORMATION FOR SEQ ID NO:29:

    (i) SEQUENCE CHARACTERISTICS:
         (A)  LENGTH:  160 base pairs
         (B)  TYPE:  nucleic acid
         (C)  STRANDEDNESS:  double
         (D)  TOPOLOGY:  linear

    (ii) MOLECULE TYPE:  DNA (genomic)
```

123

     (vi) ORIGINAL SOURCE:
         (B) STRAIN: E. coli
         (G) CELL TYPE: DH5 alpha

    (vii) IMMEDIATE SOURCE:
         (B) CLONE: C15

     (ix) FEATURE:
         (A) NAME/KEY: CDS
         (B) LOCATION: 2..151
         (D) OTHER INFORMATION: /function= "synthetic
                                   storage protein"
                                   /product= "protein"
                                   /gene= "ssp"
                                   /standard_name=
                                   "5.7.7.7.7.7.5"

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG      46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met
  1               5                   10                  15

GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG  94
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
            20                  25                  30

AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAA GAG AAG ATG 142
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met
        35                  40                  45

AAG GCG TGATAGGTAC CG                                             160
Lys Ala
        50
```

(2)   INFORMATION FOR SEQ ID NO:30:

      (i) SEQUENCE CHARACTERISTICS:
         (A)   LENGTH: 49 amino acids
         (B)   TYPE: amino acid
         (D)   TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
1               5                   10                  15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
            20                  25                  30
```

Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met Lys
        35   40   45

Ala

(2)   INFORMATION FOR SEQ ID NO:31:

            (i)   SEQUENCE CHARACTERISTICS:
                  (A) LENGTH: 160 base pairs
                  (B) TYPE: nucleic acid
                  (C) STRANDEDNESS: double
                  (D) TOPOLOGY: linear

           (ii)   MOLECULE TYPE: DNA (genomic)

           (vi)   ORIGINAL SOURCE:
                  (B) STRAIN: E. coli
                  (G) CELL TYPE: DH5 alpha

          (vii)   IMMEDIATE SOURCE:
                  (B) CLONE: C20

          (ix)   FEATURE:
                  (A) NAME/KEY: CDS
                  (B) LOCATION: 2..151
                  (D) OTHER INFORMATION: /function= "synthetic
                                           storage protein"
                                           /product= "protein"
                                           /gene= "ssp"
                                           /standard_name=
                                           "5.7.7.7.7.7.5"

          (xi)   SEQUENCE DESCRIPTION: SEQ ID NO:31:

C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG   46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met
    1   5   10   15

GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG   94
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
                20   25   30

AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAA GAG AAG ATG   142
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met
                35   40   45

AAG GCG TGATAGGTAC CG   160
Lys Ala
        50

(2)   INFORMATION FOR SEQ ID NO:32:

          (i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 49 amino acids

125

```
        (B)   TYPE:  amino acid
        (D)   TOPOLOGY:  linear

    (ii)  MOLECULE TYPE:  protein

    (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:32:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
 1               5                  10                 15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
            20                  25                 30

Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met Lys
        35                  40                 45

Ala
```

(2)   INFORMATION FOR SEQ ID NO:33:

```
        (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH:  139 base pairs
              (B)   TYPE:  nucleic acid
              (C)   STRANDEDNESS:  double
              (D)   TOPOLOGY:  linear

        (ii)  MOLECULE TYPE:  DNA (genomic)

        (vi)  ORIGINAL SOURCE:
              (B)   STRAIN:  E. coli
              (G)   CELL TYPE:  DH5 alpha

        (vii) IMMEDIATE SOURCE:
              (B)   CLONE:  C30

        (ix)  FEATURE:
              (A)   NAME/KEY:  CDS
              (B)   LOCATION:  2..130
              (D)   OTHER INFORMATION:   /function= "synthetic
                                         storage protein"
                                         /product= "protein"
                                         /gene= "ssp"
                                         /standard_name=
                                         "5.7.7.7.7.5"

        (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:33:

C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG        46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met
   1               5                  10                 15

GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG    94
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
            20                  25                 30
```

AAG CTG AAG GCG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC CG   139
Lys Leu Lys Ala Met Glu Glu Lys Met Lys Ala
            35   40

(2)   INFORMATION FOR SEQ ID NO:34:

            (i) SEQUENCE CHARACTERISTICS:
                 (A)   LENGTH: 42 amino acids
                 (B)   TYPE: amino acid
                 (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: protein

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
 1   5   10   15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
            20   25   30

Leu Lys Ala Met Glu Glu Lys Met Lys Ala
        35   40

(2)   INFORMATION FOR SEQ ID NO:35:

            (i) SEQUENCE CHARACTERISTICS:
                 (A)   LENGTH: 97 base pairs
                 (B)   TYPE: nucleic acid
                 (C)   STRANDEDNESS: double
                 (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (vi) ORIGINAL SOURCE:
                 (B)   STRAIN: E. coli
                 (G)   CELL TYPE: DH5 alpha

       (vii) IMMEDIATE SOURCE:
                 (B)   CLONE: D16

        (ix) FEATURE:
                 (A)   NAME/KEY: CDS
                 (B)   LOCATION: 2..88
                 (D)   OTHER INFORMATION: /function= "synthetic
                                            storage protein"
                                            /product= "protein"
                                            /gene= "ssp"
                                            /standard_name=
                                            "5.5.5.5"

127

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG    46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
   1               5                  10                  15


GAG GAG AAG ATG AAG GCG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC    95
Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
              20                  25


CG                                                                97
```

(2)   INFORMATION FOR SEQ ID NO:36:

      (i)   SEQUENCE CHARACTERISTICS:
          (A)   LENGTH: 28 amino acids
          (B)   TYPE: amino acid
          (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
 1               5                  10                  15


Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
              20                  25
```

(2)   INFORMATION FOR SEQ ID NO:37:

      (i)   SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 118 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (vi) ORIGINAL SOURCE:
          (B) STRAIN: E. coli
          (G) CELL TYPE: DH5 alpha

     (vii) IMMEDIATE SOURCE:
          (B) CLONE: D20

      (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 2..109
          (D) OTHER INFORMATION: /function= "synthetic
                             storage protein"
                             /product= "protein"
                             /gene= "ssp"

/standard_name=
"5.5.5.5.5"

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG      46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
   1           5               10                  15

GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG GAA GAG   94
Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu Glu
            20              25                  30

AAG ATG AAG GCG TGATAGGTAC CG                                     118
Lys Met Lys Ala
            35
```

(2)   INFORMATION FOR SEQ ID NO:38:

      (i) SEQUENCE CHARACTERISTICS:
         (A)   LENGTH: 35 amino acids
         (B)   TYPE: amino acid
         (D)   TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
 1           5               10                  15

Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys
            20              25                  30

Met Lys Ala
      35
```

(2)   INFORMATION FOR SEQ ID NO:39:

      (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 97 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double
         (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

     (vi) ORIGINAL SOURCE:
         (B) STRAIN: E. coli
         (G) CELL TYPE: DH5 alpha

    (vii) IMMEDIATE SOURCE:
         (B) CLONE: D33

```
          (ix)  FEATURE:
                (A)  NAME/KEY:  CDS
                (B)  LOCATION:  2..88
                (D)  OTHER INFORMATION:  /function= "synthetic
                                         storage protein"
                                         /product= "protein"
                                         /gene= "ssp"
                                         /standard_name=
                                         "5.5.5.5"
```

```
          (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:39:
```

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG      46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
   1               5                  10                  15
```

```
GAG GAG AAG ATG AAG GCG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC     95
Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
                20                  25
```

```
CG                                                                 97
```

```
(2)    INFORMATION FOR SEQ ID NO:40:
```

```
       (i)   SEQUENCE CHARACTERISTICS:
             (A)   LENGTH:  28 amino acids
             (B)   TYPE:  amino acid
             (D)   TOPOLOGY:  linear
```

```
       (ii)  MOLECULE TYPE:  protein
```

```
       (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:40:
```

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
 1               5                  10                  15
```

```
Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
                20                  25
```

```
(2)    INFORMATION FOR SEQ ID NO:41:
```

```
       (i)   SEQUENCE CHARACTERISTICS:
             (A)   LENGTH:  21 base pairs
             (B)   TYPE:  nucleic acid
             (C)   STRANDEDNESS:  single
             (D)   TOPOLOGY:  linear
```

```
       (ii)  MOLECULE TYPE:  DNA (genomic)
```

```
       (ix)  FEATURE:
             (A)   NAME/KEY:  misc_feature
             (B)   LOCATION:  1..21
             (D)   OTHER INFORMATION:  /product= "synthetic
                                       oligonucleotide"
```

                                          /standard_name= "SM
                                          86"

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATGGAGGAG AAGCTGAAGA A   21

   (2)   INFORMATION FOR SEQ ID NO:42:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 21 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (ix) FEATURE:
             (A)   NAME/KEY: misc_feature
             (B)   LOCATION: 1..21
             (D)   OTHER INFORMATION: /product= "synthetic
                                        oligonucleotide"
                                        /standard_name= "SM
                                        87"

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ATCTTCTTCA GCTTCTCCTC C   21

   (2)   INFORMATION FOR SEQ ID NO:43:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 21 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (ix) FEATURE:
             (A)   NAME/KEY: misc_feature
             (B)   LOCATION: 1..21
             (D)   OTHER INFORMATION: /product= "synthetic
                                        oligonucleotide"
                                        /standard_name= "SM
                                        88"

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GATGGAGGAG AAGCTGAAGT G   21

131

(2)   INFORMATION FOR SEQ ID NO:44:

    (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 21 base pairs
        (B)   TYPE: nucleic acid
        (C)   STRANDEDNESS: single
        (D)   TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

   (ix) FEATURE:
        (A)   NAME/KEY: misc_feature
        (B)   LOCATION: 1..21
        (D)   OTHER INFORMATION: /product= "synthetic
                                  oligonucleotide"
                                  /standard_name= "SM
                                  89"

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATCCACTTCA GCTTCTCCTC C   21

(2)   INFORMATION FOR SEQ ID NO:45:

    (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 21 base pairs
        (B)   TYPE: nucleic acid
        (C)   STRANDEDNESS: single
        (D)   TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

   (ix) FEATURE:
        (A)   NAME/KEY: misc_feature
        (B)   LOCATION: 1..21
        (D)   OTHER INFORMATION: /product= "synthetic
                                  oligonucleotide"
                                  /standard_name= "SM
                                  90"

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GATGGAGGAG AAGATGAAGA A   21

(2)   INFORMATION FOR SEQ ID NO:46:

    (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 21 base pairs
        (B)   TYPE: nucleic acid
        (C)   STRANDEDNESS: single
        (D)   TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA (genomic)

132

```
      (ix) FEATURE:
           (A)   NAME/KEY:  misc_feature
           (B)   LOCATION:  1..21
           (D)   OTHER INFORMATION:  /product= "synthetic
                                     oligonucleotide"
                                     /standard_name= "SM
                                     91"


      (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:46:

ATCTTCTTCA TCTTCTCCTC C                                        21

 (2)   INFORMATION FOR SEQ ID NO:47:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  21 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

       (ii) MOLECULE TYPE:  DNA (genomic)

       (ix) FEATURE:
            (A)   NAME/KEY:  misc_feature
            (B)   LOCATION:  1..21
            (D)   OTHER INFORMATION:  /product= "synthetic
                                      oligonucleotide"
                                      /standard_name= "SM
                                      92"


      (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:47:

GATGGAGGAG AAGATGAAGT G                                        21

 (2)   INFORMATION FOR SEQ ID NO:48:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  21 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

       (ii) MOLECULE TYPE:  DNA (genomic)

       (ix) FEATURE:
            (A)   NAME/KEY:  misc_feature
            (B)   LOCATION:  1..21
            (D)   OTHER INFORMATION:  /product= "synthetic
                                      oligonucleotide"
                                      /standard_name= "SM
                                      93"
```

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

ATCCACTTCA TCTTCTCCTC C   21

  (2)   INFORMATION FOR SEQ ID NO:49:

      (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 7 amino acids
        (B)   TYPE: amino acid
        (C)   STRANDEDNESS: unknown
        (D)   TOPOLOGY: unknown

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Met Glu Glu Lys Leu Lys Lys
1   5

  (2)   INFORMATION FOR SEQ ID NO:50:

      (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 7 amino acids
        (B)   TYPE: amino acid
        (C)   STRANDEDNESS: unknown
        (D)   TOPOLOGY: unknown

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met Glu Glu Lys Leu Lys Trp
1   5

  (2)   INFORMATION FOR SEQ ID NO:51:

      (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 7 amino acids
        (B)   TYPE: amino acid
        (C)   STRANDEDNESS: unknown
        (D)   TOPOLOGY: unknown

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met Glu Glu Lys Met Lys Lys
1   5

  (2)   INFORMATION FOR SEQ ID NO:52:

      (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 7 amino acids

```
        (B)   TYPE:  amino acid
        (C)   STRANDEDNESS:  unknown
        (D)   TOPOLOGY:  unknown

   (ii) MOLECULE TYPE:  protein

   (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:52:

Met Glu Glu Lys Met Lys Trp
1                   5

(2)  INFORMATION FOR SEQ ID NO:53:

        (i)   SEQUENCE CHARACTERISTICS:
              (A)   LENGTH:  160 base pairs
              (B)   TYPE:  nucleic acid
              (C)   STRANDEDNESS:  double
              (D)   TOPOLOGY:  linear

       (ii) MOLECULE TYPE:  DNA (genomic)

       (vi) ORIGINAL SOURCE:
              (B)  STRAIN:  E. coli
              (G)  CELL TYPE:  DH5 alpha

      (vii) IMMEDIATE SOURCE:
              (B)  CLONE:  82-4

      (ix) FEATURE:
              (A)  NAME/KEY:  CDS
              (B)  LOCATION:  2..151
              (D)  OTHER INFORMATION:   /function= "synthetic
                                        storage protein
                                        /product= "protein"
                                        /gene= "ssp"
                                        /standard_name=
                                        "7.7.7.7.7.7.5"

       (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:53:

C ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG      46
  Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met
   1           5                   10                      15

GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG  94
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
            20                  25                      30

AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAA GAG AAG ATG 142
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met
            35                  40                      45
```

135

AAG GCG TGATAGGTAC CG   160
Lys Ala
        50

(2)   INFORMATION FOR SEQ ID NO:54:

       (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 49 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
 1   5   10   15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
            20   25   30

Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met Lys
        35   40   45

Ala

(2)   INFORMATION FOR SEQ ID NO:55:

       (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 97 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (vi) ORIGINAL SOURCE:
            (B) STRAIN: E. coli
            (G) CELL TYPE: DH5 alpha

       (vii) IMMEDIATE SOURCE:
            (B) CLONE: 84-H3

       (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION: 2..88
            (D) OTHER INFORMATION: /function= "synthetic
                                    storage protein
                                    /product= "protein"
                                    /gene= "ssp"
                                    /standard_name=
                                    "5.5.5.5"

136

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG    46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
   1               5                   10                  15

GAG GAG AAG ATG AAG GCG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC   95
Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
             20                  25

CG                                                               97
```

(2)   INFORMATION FOR SEQ ID NO:56:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 28 amino acids
             (B)   TYPE: amino acid
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
 1               5                   10                  15

Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
         20                  25
```

(2)   INFORMATION FOR SEQ ID NO:57:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 97 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: double
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (vi) ORIGINAL SOURCE:
             (B)   STRAIN: E. coli
             (G)   CELL TYPE: DH5 alpha

      (vii) IMMEDIATE SOURCE:
             (B)   CLONE: 86-H23

      (ix)   FEATURE:
             (A)   NAME/KEY: CDS
             (B)   LOCATION: 2..88
             (D)   OTHER INFORMATION: /function= "synthetic
                                       storage protein
                                       /product= "protein"
                                       /gene= "ssp"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG CTG AAG AAG ATG     46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Lys Met
   1               5                   10                  15

GAG GAG AAG CTG AAG AAG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC    95
Glu Glu Lys Leu Lys Lys Met Glu Glu Lys Met Lys Ala
             20                  25

CG                                                               97
```

(2)   INFORMATION FOR SEQ ID NO:58:

       (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 28 amino acids
            (B)   TYPE: amino acid
            (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```
Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Lys Met Glu
  1               5                   10                  15

Glu Lys Leu Lys Lys Met Glu Glu Lys Met Lys Ala
             20                  25
```

(2)   INFORMATION FOR SEQ ID NO:59:

       (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 112 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (vi) ORIGINAL SOURCE:
            (B) STRAIN: E. coli
            (G) CELL TYPE: DH5 alpha

     (vii) IMMEDIATE SOURCE:
            (B) CLONE: 88-2

      (ix) FEATURE:
            (A) NAME/KEY: CDS
            (B) LOCATION: 2..103
            (D) OTHER INFORMATION: /function= "synthetic
                                    storage protein

```
                                    /product= "protein"
                                    /gene= "ssp"
                                    /standard_name=
                                    "5.9.9.9.5"
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```
C ATG GAG GAG AAG ATG AAG GCG AAG AAG CTG AAG TGG ATG GAG GAG    46
  Met Glu Glu Lys Met Lys Ala Lys Lys Leu Lys Trp Met Glu Glu
   1               5                  10                  15

AAG CTG AAG TGG ATG GAG GAG AAG CTG AAG TGG ATG GAA GAG AAG ATG 94
Lys Leu Lys Trp Met Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Met
             20           .        25                  30

AAG GCG TGATAGGTAC CG                                           112
Lys Ala
```

(2)   INFORMATION FOR SEQ ID NO:60:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 33 amino acids
             (B)   TYPE: amino acid
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

```
Met Glu Glu Lys Met Lys Ala Lys Lys Leu Lys Trp Met Glu Glu Lys
 1               5                  10                  15

Leu Lys Trp Met Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Met Lys
             20                  25                  30

Ala
```

(2)   INFORMATION FOR SEQ ID NO:61:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 118 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: double
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (vi) ORIGINAL SOURCE:
             (B) STRAIN: E. coli
             (G) CELL TYPE: DH5 alpha

      (vii) IMMEDIATE SOURCE:
             (B) CLONE: 90-H8

139

```
          (ix)  FEATURE:
                (A)  NAME/KEY:  CDS
                (B)  LOCATION:  2..109
                (D)  OTHER INFORMATION:  /function= "synthetic
                                         storage protein
                                         /product= "protein"
                                         ./gene= "ssp"
                                         /standard_name=
                                         "5.10.10.10.5"

          (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:61:

C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG AAG ATG    46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Lys Met
   1           5                    10                   15

GAG GAG AAG ATG AAG AAG ATG GAG GAG AAG ATG AAG AAG ATG GAA GAG 94
Glu Glu Lys Met Lys Lys Met Glu Glu Lys Met Lys Lys Met Glu Glu
            20                   25                   30

AAG ATG AAG GCG TGATAGGTAC CG                                    118
Lys Met Lys Ala
            35
```

(2)   INFORMATION FOR SEQ ID NO:62:

```
          (i)  SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 35 amino acids
                (B)   TYPE: amino acid
                (D)   TOPOLOGY: linear

         (ii)  MOLECULE TYPE:  protein

         (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:62:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Lys Met Glu
 1           5                    10                   15

Glu Lys Met Lys Lys Met Glu Glu Lys Met Lys Lys Met Glu Glu Lys
            20                   25                   30

Met Lys Ala
        35
```

(2)   INFORMATION FOR SEQ ID NO:63:

```
          (i)  SEQUENCE CHARACTERISTICS:
                (A)   LENGTH:  97 base pairs
                (B)   TYPE:  nucleic acid
                (C)   STRANDEDNESS:  double
                (D)   TOPOLOGY:  linear

         (ii)  MOLECULE TYPE:  DNA (genomic)
```

140

```
        (vi)  ORIGINAL SOURCE:
              (B)  STRAIN:  E. coli
              (G)  CELL TYPE:  DH5 alpha

       (vii)  IMMEDIATE SOURCE:
              (B)  CLONE:  92-2

        (ix)  FEATURE:
              (A)  NAME/KEY:  CDS
              (B)  LOCATION:  2..88
              (D)  OTHER INFORMATION:  /function= "synthetic
                                      storage protein
                                      /product= "protein"
                                      /gene= "ssp"
                                      /standard_name=
                                      "5.11.11.5"

        (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:63:

C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG TGG ATG        46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Trp Met
   1               5                  10                  15

GAG GAG AAG ATG AAG TGG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC      95
Glu Glu Lys Met Lys Trp Met Glu Glu Lys Met Lys Ala
              20                  25

CG                                                                  97

  (2)    INFORMATION FOR SEQ ID NO:64:

         (i)  SEQUENCE CHARACTERISTICS:
              (A)    LENGTH:  28 amino acids
              (B)    TYPE:  amino acid
              (D)    TOPOLOGY:  linear

        (ii)  MOLECULE TYPE:  protein

        (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:64:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Trp Met Glu
 1               5                  10                  15

Glu Lys Met Lys Trp Met Glu Glu Lys Met Lys Ala
            20                  25

  (2)    INFORMATION FOR SEQ ID NO:65:

         (i)  SEQUENCE CHARACTERISTICS:
              (A)    LENGTH:  84 base pairs
              (B)    TYPE:  nucleic acid
              (C)    STRANDEDNESS:  single
              (D)    TOPOLOGY:  linear
```

141

        (ii) MOLECULE TYPE: DNA (genomic)

        (ix) FEATURE:
              (A)   NAME/KEY: misc_feature
              (B)   LOCATION: 1..84
              (D)   OTHER INFORMATION: /product= "synthetic
                                        ·oligonucleotide"
                                        /standard_name= "SM
                                        96"

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GATGGAGGAA AAGATGAAGG CGATGGAGGA GAAAATGAAA GCTATGGAGG AAAAGATGAA 60

AGCGATGGAG GAGAAAATGA AGGC   84

    (2)   INFORMATION FOR SEQ ID NO:66:

          (i) SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 84 base pairs
               (B)   TYPE: nucleic acid
               (C)   STRANDEDNESS: single
               (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (ix) FEATURE:
              (A)   NAME/KEY: misc_feature
              (B)   LOCATION: 1..84
              (D)   OTHER INFORMATION: /product= "synthetic
                                        oligonucleotide"
                                        /standard_name= "SM
                                        97"

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ATCGCCTTCA TTTTCTCCTC CATCGCTTTC ATCTTTTCCT CCATAGCTTT CATTTTCTCC 60

TCCATCGCCT TCATCTTTTC CTCC   84

    (2)   INFORMATION FOR SEQ ID NO:67:

          (i) SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 28 amino acids
               (B)   TYPE: amino acid
               (C)   STRANDEDNESS: unknown
               (D)   TOPOLOGY: unknown

        (ii) MOLECULE TYPE: protein

        (ix) FEATURE:
              (A)   NAME/KEY: Protein
              (B)   LOCATION: 1..28

            (D)   OTHER INFORMATION: /label= name
                                      /note= "(SSP 5)4"

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
1   5   10   15

Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
            20   25

   (2)   INFORMATION FOR SEQ ID NO:68:

         (i)   SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 84 base pairs
               (B)   TYPE: nucleic acid
               (C)   STRANDEDNESS: single
               (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (ix) FEATURE:
               (A)   NAME/KEY: misc_feature
               (B)   LOCATION: 1..84
               (D)   OTHER INFORMATION: /product= "synthetic
                                          oligonucleotide"
                                          /standard_name= "SM
                                          98"

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GATGGAGGAA AAGCTGAAAG CGATGGAGGA GAAACTCAAG GCTATGGAAG AAAAGCTTAA 60

AGCGATGGAG GAGAAACTGA AGGC   84

   (2)   INFORMATION FOR SEQ ID NO:69:

         (i)   SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 84 base pairs
               (B)   TYPE: nucleic acid
               (C)   STRANDEDNESS: single
               (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (ix) FEATURE:
               (A)   NAME/KEY: misc_feature
               (B)   LOCATION: 1..84
               (D)   OTHER INFORMATION: /product= "synthetic
                                          oligonucleotide"
                                          /standard_name= "SM
                                          99"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATCGCCTTCA GTTTCTCCTC CTACGCTTTA AGCTTTTCTT CCATAGCCTT GAGTTTCTCC 60

TCCATCGCTT TCAGCTTTTC CTCC   84

(2) INFORMATION FOR SEQ ID NO:70:

      (i) SEQUENCE CHARACTERISTICS:
          (A)   LENGTH: 28 amino acids
          (B)   TYPE: amino acid
          (C)   STRANDEDNESS: unknown
          (D)   TOPOLOGY: unknown

      (ii) MOLECULE TYPE: protein

      (ix) FEATURE:
          (A)   NAME/KEY: Protein
          (B)   LOCATION: 1..28
          (D)   OTHER INFORMATION: /label= name
                                       /note= "(SSP 7)4"

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
1   5   10   15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala
        20   25

(2) INFORMATION FOR SEQ ID NO:71:

      (i) SEQUENCE CHARACTERISTICS:
          (A)   LENGTH: 84 base pairs
          (B)   TYPE: nucleic acid
          (C)   STRANDEDNESS: single
          (D)   TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (ix) FEATURE:
          (A)   NAME/KEY: misc_feature
          (B)   LOCATION: 1..84
          (D)   OTHER INFORMATION: /product= "synthetic
                                       oligonucleotide"
                                       /standard_name= "SM
                                       100"

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GATGGAGGAA AAGCTTAAGA AGATGGAAGA AAAGCTGAAA TGGATGGAGG AGAAACTCAA 60

AAAGATGGAG GAAAAGCTTA AATG   84

(2)   INFORMATION FOR SEQ ID NO:72:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 84 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
             (A) NAME/KEY: misc_feature
             (B) LOCATION: 1..84
             (D) OTHER INFORMATION: /product= "synthetic
                                     oligonucleotide"
                                     /standard_name= "SM
                                     101"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

ATCCATTTAA GCTTTTCCTC CTACTTTTTG AGTTTCTCCT CCATCCATTT CAGCTTTTCT 60

TCCATCTTCT TAAGCTTTTC CTCC   84

(2)   INFORMATION FOR SEQ ID NO:73:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 28 amino acids
             (B) TYPE: amino acid
             (C) STRANDEDNESS: unknown
             (D) TOPOLOGY: unknown

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Met Glu Glu Lys Leu Lys Lys Met Glu Glu Lys Leu Lys Trp Met Glu
1   5   10   15

Glu Lys Leu Lys Lys Met Glu Glu Lys Leu Lys Trp
            20   25

(2)   INFORMATION FOR SEQ ID NO:74:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 243 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: double
             (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (vi) ORIGINAL SOURCE:
             (B) STRAIN: E. coli

145

              (G) CELL TYPE: DH5 alpha

     (vii)   IMMEDIATE SOURCE:
             (B) CLONE: 2-9

      (ix)   FEATURE:
             (A) NAME/KEY: CDS
             (B) LOCATION: 2..235
             (D) OTHER INFORMATION: /function= "synthetic
                                     storage protein
                                     /product= "protein"
                                     /gene= "ssp"
                                     /standard_name=
                                     "7.7.7.7.7.7.8.9.8.9.5"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```
C ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG        46
  Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met
   1               5               10               15

GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG      94
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
             20               25               30

AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAA AAG CTT     142
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu
         35               40               45

AAG AAG ATG GAA GAA AAG CTG AAA TGG ATG GAG GAG AAA CTC AAA AAG     190
Lys Lys Met Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Leu Lys Lys
     50               55               60

ATG GAG GAA AAG CTT AAA TGG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC  242
Met Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Met Lys Ala
 65               70               75

C                                                                  243
```

(2)   INFORMATION FOR SEQ ID NO:75:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 77 amino acids
             (B)   TYPE: amino acid
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

```
Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
 1               5               10               15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
             20               25               30
```

```
Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys
        35              40              45

Lys Met Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Leu Lys Lys Met
    50              55              60

Glu Glu Lys Leu Lys Trp Met Glu Glu Lys Met Lys Ala
65              70              75
```

(2)   INFORMATION FOR SEQ ID NO:76:

          (i)   SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 175 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: double
                (D) TOPOLOGY: linear

         (ii)   MOLECULE TYPE: DNA (genomic)

         (vi)   ORIGINAL SOURCE:
                (B) STRAIN: E. coli
                (G) CELL TYPE: DH5 alpha

        (vii)   IMMEDIATE SOURCE:
                (B) CLONE: 5-1

         (ix)   FEATURE:
                (A) NAME/KEY: CDS
                (B) LOCATION: 2..172
                (D) OTHER INFORMATION: /function= "synthetic
                                           storage protein
                                           /product= "protein"
                                           /gene= "ssp"
                                           /standard_name=
                                           "5.5.5.7.7.7.7.5"

         (xi)   SEQUENCE DESCRIPTION: SEQ ID NO:76:

```
C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG     46
  Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
   1           5               10                  15

GAG GAG AAG ATG AAG GCG ATG GAG GAA AAG CTG AAA GCG ATG GAG GAG  94
Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
            20              25                  30

AAA CTC AAG GCT ATG GAA GAA AAG CTT AAA GCG ATG GAG GAG AAA CTG 142
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu
            35              40                  45

AAG GCC ATG GAA GAG AAG ATG AAG GCG TGATAG                       179
Lys Ala Met Glu Glu Lys Met Lys Ala
            50              55
```

147

(2)   INFORMATION FOR SEQ ID NO:77:

          (i) SEQUENCE CHARACTERISTICS:
               (A) LENGTH: 56 amino acids
               (B) TYPE: amino acid
               (D) TOPOLOGY: linear

         (ii) MOLECULE TYPE: protein

         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
 1   5   10   15

Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
          20   25   30

Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys
     35   40   45

Ala Met Glu Glu Lys Met Lys Ala
     50   55

(2)   INFORMATION FOR SEQ ID NO:78:

          (i) SEQUENCE CHARACTERISTICS:
               (A) LENGTH: 187 base pairs
               (B) TYPE: nucleic acid
               (C) STRANDEDNESS: double
               (D) TOPOLOGY: linear

         (ii) MOLECULE TYPE: DNA (genomic)

         (vi) ORIGINAL SOURCE:
               (B) STRAIN: E. coli
               (G) CELL TYPE: DH5 alpha

         (ix) FEATURE:
               (A) NAME/KEY: CDS
               (B) LOCATION: 3..173
               (D) OTHER INFORMATION: /function= "synthetic
                                        storage protein
                                        /product= "protein"
                                        /gene= "ssp"
                                        /standard_name=
                                        "SSP-3-5"

         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CC ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG   47
   Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met
    1   5   10   15

148

```
GAG GAG AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAG    95
Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
            20              25                  30

AAG CTG AAG GCG ATG GAG GAG AAG CTG AAG GCG ATG GAG GAA AAG ATG   143
Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met
        35              40                  45

AAG GCG ATG GAA GAG AAG ATG AAG GCG TGATAGGTAC CGAATTC           187
Lys Ala Met Glu Glu Lys Met Lys Ala
        50              55
```

(2)   INFORMATION FOR SEQ ID NO:79:

          (i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 56 amino acids
                (B)   TYPE: amino acid
                (D)   TOPOLOGY: linear

         (ii)   MOLECULE TYPE: protein

         (xi)   SEQUENCE DESCRIPTION: SEQ ID NO:79:

```
Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu
 1               5              10                  15

Glu Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys
            20              25                  30

Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Met Lys
        35              40                  45

Ala Met Glu Glu Lys Met Lys Ala
        50              55
```

(2)   INFORMATION FOR SEQ ID NO:80:

          (i)   SEQUENCE CHARACTERISTICS:
                (A)   LENGTH: 61 base pairs
                (B)   TYPE: nucleic acid
                (C)   STRANDEDNESS: single
                (D)   TOPOLOGY: linear

         (ii)   MOLECULE TYPE: DNA (genomic)

         (ix)   FEATURE:
                (A)   NAME/KEY: misc_feature
                (B)   LOCATION: 1..61
                (D)   OTHER INFORMATION: /product= "synthetic
                                           oligonucleotide"
                                           /standard_name= "SM
                                           107"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

CATGGAGGAG AAGATGAAAA AGCTCGAAGA GAAGATGAAG GTCATGAAGT GATAGGTACC 60

G   61

(2)   INFORMATION FOR SEQ ID NO:81:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 61 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
             (A)   NAME/KEY: misc_feature
             (B)   LOCATION: 1..61
             (D)   OTHER INFORMATION: /product= "synthetic
                                      oligonucleotide"
                                      /standard_name= "SM
                                      106"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

AATTCGGTAC CTATCACTTC ATGACCTTCA TCTTCTCTTC GAGCTTTTTC ATCTTCTCCT 60

C   61

(2)   INFORMATION FOR SEQ ID NO:82:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 16 amino acids
             (B)   TYPE: amino acid
             (C)   STRANDEDNESS: unknown
             (D)   TOPOLOGY: unknown

       (ii) MOLECULE TYPE: protein

       (ix) FEATURE:
             (A)   NAME/KEY: Protein
             (B)   LOCATION: 1..16
             (D)   OTHER INFORMATION: /label= name
                                      /note= "pSK34 base
                                      gene"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met Glu Glu Lys Met Lys Lys Leu Glu Glu Lys Met Lys Val Met Lys
1   5   10   15

150

(2)   INFORMATION FOR SEQ ID NO:83:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 63 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
             (A)   NAME/KEY: misc_feature
             (B)   LOCATION: 1..63
             (D)   OTHER INFORMATION: /product= "synthetic
                                        oligonucleotide"
                                        /standard_name= "SM
                                        110"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GCTGGAAGAA AAGATGAAGG CTATGGAGGA CAAGATGAAA TGGCTTGAGG AAAAGATGAA 60

GAA   63

(2)   INFORMATION FOR SEQ ID NO:84:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 63 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (ix) FEATURE:
             (A)   NAME/KEY: misc_feature
             (B)   LOCATION: 1..63
             (D)   OTHER INFORMATION: /product= "synthetic
                                        oligonucleotide"
                                        /standard_name= "SM
                                        111"

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

AGCTTCTTCA TCTTTTCCTC AAGCCATTTC ATCTTGTCCT CCATAGCCTT CATCTTTTCT 60

TCC   63

(2)   INFORMATION FOR SEQ ID NO:85:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 37 amino acids
             (B)   TYPE: amino acid
             (D)   TOPOLOGY: linear

151

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```
Met Glu Glu Lys Met Lys Lys Leu Glu Glu Lys Met Lys Ala Met Glu
 1               5              10                  15

Asp Lys Met Lys Trp Leu Glu Glu Lys Met Lys Lys Leu Glu Glu Lys
            20              25                  30

Met Lys Val Met Lys
        35
```

(2)   INFORMATION FOR SEQ ID NO:86:

    (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 37 amino acids
        (B)   TYPE: amino acid
        (D)   TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

```
Met Glu Glu Lys Met Lys Lys Leu Glu Glu Lys Met Lys Ala Met Glu
 1               5              10                  15

Asp Lys Met Lys Trp Leu Glu Glu Lys Met Lys Lys Leu Glu Glu Lys
            20              25                  30

Met Lys Val Met Lys
        35
```

(2)   INFORMATION FOR SEQ ID NO:87:

    (i) SEQUENCE CHARACTERISTICS:
        (A)   LENGTH: 62 base pairs
        (B)   TYPE: nucleic acid
        (C)   STRANDEDNESS: single
        (D)   TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (ix) FEATURE:
        (A)   NAME/KEY: misc_feature
        (B)   LOCATION: 1..62
        (D)   OTHER INFORMATION: /product= "synthetic
                             oligonucletide"
                             /standard_name= "SM
                             112"

152

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GCTCGAAGAA AGATGAAGGC AATGGAAGAC AAAATGAAGT GGCTTGAGGA GAAAATGAAG 60

AA   62

(2)   INFORMATION FOR SEQ ID NO:88:

          (i) SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 62 base pairs
               (B)   TYPE: nucleic acid
               (C)   STRANDEDNESS: single
               (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: DNA (genomic)

          (ix) FEATURE:
               (A)   NAME/KEY: misc_feature
               (B)   LOCATION: 1..62
               (D)   OTHER INFORMATION: /product= "synthetic
                                           oligonucleotide"
                                           /standard_name= "SM
                                           113"

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

AGCTTCTTCA TTTTCTCCTC AAGCCACTTC ATTTTGTCTT CCATTGCCTT CATCTTTCTT 60

CG   62

(2)   INFORMATION FOR SEQ ID NO:89:

          (i) SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 37 amino acids
               (B)   TYPE: amino acid
               (D)   TOPOLOGY: linear

          (ii) MOLECULE TYPE: protein

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Glu Glu Lys Met Lys Lys Leu Lys Glu Glu Met Ala Lys Met Lys
 1   5   10   15

Asp Glu Met Trp Lys Leu Lys Glu Glu Met Lys Lys Leu Glu Glu Lys
            20   25   30

Met Lys Val Met Lys
          35

(2)   INFORMATION FOR SEQ ID NO:90:

          (i) SEQUENCE CHARACTERISTICS:
               (A)   LENGTH: 63 base pairs

```
          (B)   TYPE:  nucleic acid
          (C)   STRANDEDNESS:  single
          (D)   TOPOLOGY:  linear

     (ii) MOLECULE TYPE:  DNA (genomic)

     (ix) FEATURE:
          (A)   NAME/KEY:  misc_feature
          (B)   LOCATION:  1..63
          (D)   OTHER INFORMATION:  /product= "synthetic
                                    oligonucleotide"
                                    /standard_name= "SM
                                    114"

     (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:90:
```

GCTCAAGGAG GAAATGGCTA AGATGAAAGA CGAAATCTGG AAACTGAAAG AGGAAATGAA 60

GAA   63

```
(2)     INFORMATION FOR SEQ ID NO:91:

     (i) SEQUENCE CHARACTERISTICS:
          (A)   LENGTH:  63 base pairs
          (B)   TYPE:  nucleic acid
          (C)   STRANDEDNESS:  single
          (D)   TOPOLOGY:  linear

     (ii) MOLECULE TYPE:  DNA (genomic)

     (ix) FEATURE:
          (A)   NAME/KEY:  misc_feature
          (B)   LOCATION:  1..63
          (D)   OTHER INFORMATION:  /product= "synthetic
                                    oligonucleotide"
                                    /standard_name= "SM
                                    115"

     (xi) SEQUENCE DESCRIPTION:  SEQ ID NO:91:
```

AGCTTCTTCA TTTCCTCTTT CAGTTTCCAC ATTTCGTCTT TCATCTTAGC CATTTCCTCC 60

TTG   63

```
(2)     INFORMATION FOR SEQ ID NO:92:

     (i) SEQUENCE CHARACTERISTICS:
          (A)   LENGTH:  107 amino acids
          (B)   TYPE:  amino acid
          (D)   TOPOLOGY:  linear

     (ii) MOLECULE TYPE:  protein
```

154

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Met Glu Glu Lys Met Lys Lys Leu Lys Glu Glu Met Ala Lys Met Lys
 1   5   10   15

Asp Glu Met Trp Lys Leu Lys Glu Glu Met Lys Lys Leu Glu Glu Lys
            20   25   30

Met Lys Val Met Glu Glu Lys Met Lys Lys Leu Glu Glu Lys Met Lys
        35   40   45

Ala Met Glu Asp Lys Met Lys Trp Leu Glu Glu Lys Met Lys Lys Leu
    50   55   60

Glu Glu Lys Met Lys Val Met Glu Glu Lys Met Lys Lys Leu Glu Glu
65   70   75   80

Lys Met Lys Ala Met Glu Asp Lys Met Lys Trp Leu Glu Glu Lys Met
                85   90   95

Lys Lys Leu Glu Glu Lys Met Lys Val Met Lys
            100   105

(2)   INFORMATION FOR SEQ ID NO:93:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 43 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CTAGAAGCCT CGGCAACGTC AGCAACGGCG GAAGAATCCG GTG   43

(2)   INFORMATION FOR SEQ ID NO:94:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 43 base pairs
            (B)   TYPE: nucleic acid
            (C)   STRANDEDNESS: single
            (D)   TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

CATGCACCGG ATTCTTCCGC CGTTGCTGAC GTTGCCGAGG CTT   43

(2)   INFORMATION FOR SEQ ID NO:95:

        (i) SEQUENCE CHARACTERISTICS:
            (A)   LENGTH: 55 base pairs

155

```
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

      (ii)  MOLECULE TYPE:  DNA (genomic)

      (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:95:

GATCCCATGG CGCCCCTTAA GTCCACCGCC AGCCTCCCCG TCGCCCGCCG CTCCT      55

   (2)   INFORMATION FOR SEQ ID NO:96:

          (i)  SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  55 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

      (ii)  MOLECULE TYPE:  DNA (genomic)

      (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:96:

CTAGAGGAGC GGCGGGCGAC GGGGAGGCTG GCGGTGGACT TAAGGGGCGC CATGG      55

   (2)   INFORMATION FOR SEQ ID NO:97:

          (i)  SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  59 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

      (ii)  MOLECULE TYPE:  DNA (genomic)

      (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:97:

CATGGCGCCC ACCGTGATGA TGGCCTCGTC GGCCACCGCC GTCGCTCCGT TCCAGGGGC  59

   (2)   INFORMATION FOR SEQ ID NO:98:

          (i)  SEQUENCE CHARACTERISTICS:
            (A)   LENGTH:  59 base pairs
            (B)   TYPE:  nucleic acid
            (C)   STRANDEDNESS:  single
            (D)   TOPOLOGY:  linear

      (ii)  MOLECULE TYPE:  DNA (genomic)

      (xi)  SEQUENCE DESCRIPTION:  SEQ ID NO:98:

TTAAGCCCCT GGAACGGAGC GACGGCGGTG GCCGACGAGG CCATCATCAC GGTGGGCGC  59
```

156

(2)   INFORMATION FOR SEQ ID NO:99:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 16 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GCGCCCACCG TGATGA   16

(2)   INFORMATION FOR SEQ ID NO:100:

        (i) SEQUENCE CHARACTERISTICS:
             (A)   LENGTH: 16 base pairs
             (B)   TYPE: nucleic acid
             (C)   STRANDEDNESS: single
             (D)   TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (genomic)

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

CACCGGATTC TTCCGC   16

<u>CLAIMS</u>

What is claimed is:

1.   A chimeric gene wherein a nucleic acid
fragment encoding dihydrodipicolinic acid synthase which
is insensitive to inhibition by lysine is operably
linked to a plant chloroplast transit sequence and to a
plant seed-specific regulatory sequence.

2.   The chimeric gene of Claim 1, wherein the
nucleic acid fragment encoding dihydrodipicolinic acid
synthase comprises the nucleotide sequence shown in SEQ
ID NO:3: encoding dihydrodipicolinic acid synthase from
<u>Corynebacterium glutamicum</u> and wherein the plant
chloroplast transit sequence is derived from a gene
encoding the small subunit of ribulose 1,5-bisphosphate
carboxylase from <u>Glycine max</u>, and the seed-specific
regulatory sequence is from the gene encoding the β
subunit of the seed storage protein phaseolin from the
bean <u>Phaseolus vulgaris</u> or the seed-specific regulatory
sequence is from the Kunitz trypsin inhibitor 3 gene
from <u>Glycine max</u>.

3.   A plant comprising in its genome the chimeric
gene of Claim 1 or Claim 2.

4.   Seed obtained from the plant of Claim 3.

5.   A method for obtaining a plant wherein the
seeds of the plant accumulate lysine at a level from ten
percent to four hundred percent higher than do seeds of
an untransformed plant comprising:

     (a)   transforming plant cells with the
chimeric gene of Claim 1;

     (b)   regenerating fertile mature plants from
the transformed plant cells obtained from step (a) under
conditions suitable to obtain seeds;

     (c)   screening the progeny seed of step (b)
for lysine content; and

(d)   selecting those lines whose seeds contain increased levels of lysine.

6.   A method for obtaining a dicot plant wherein the seeds of the plant accumulate lysine at a level from ten percent to one hundred percent higher than do seeds of an untransformed plant comprising:

(a)   transforming dicot cells with the chimeric gene of Claim 1 or Claim 2;

(b)   regenerating fertile mature plants from the transformed plant cells obtained from step (a) under conditions suitable to obtain seeds;

(c)   screening the progeny seed of step (b) for lysine content; and

(d)   selecting those lines whose seeds contain increased levels of lysine.

7.   A method for obtaining a rapeseed plant wherein the seeds of the plant accumulate lysine at a level from ten percent to one hundred percent higher than do seeds of an untransformed plant comprising:

(a)   transforming rapeseed cells with the chimeric gene of Claim 1 or Claim 2;

(b)   regenerating fertile mature plants from the transformed plant cells obtained from step (a) under conditions suitable to obtain seeds;

(c)   screening the progeny seed of step (b) for lysine content; and

(d)   selecting those lines whose seeds contain increased levels of lysine.

8.   A method for obtaining a soybean plant wherein the seeds of the plant accumulate lysine at a level from ten percent to four hundred percent higher than do seeds of an untransformed plant comprising:

(a)   transforming soybean cells with the chimeric gene of Claim 1 or Claim 2;

        (b)   regenerating fertile mature plants from the transformed plant cells obtained from step (a) under conditions suitable to obtain seeds;

        (c)   screening the progeny seed of step (b)
5   for lysine content; and

        (d)   selecting those lines whose seeds contain increased levels of lysine.

        9.   A transformed plant wherein the seeds of the plant accumulate lysine at a level at least ten percent
10 higher than do seeds of an untransformed plant.

        10.   A transformed plant, as described by Claim 9, wherein the seeds of the plant accumulate lysine at a level from ten percent to four hundred percent higher than do seeds of an untransformed plant.

15   11.   A transformed rapeseed plant wherein the seeds of the plant accumulate lysine to a level between ten percent and one hundred percent higher than do seeds of an untransformed plant.

        12.   A transformed soybean plant wherein the seeds
20 of the plant accumulate lysine to a level between ten percent and four hundred percent higher than do seeds of an untransformed plant.

        13.   A chimeric gene of Claim 1 wherein the seed-specific regulatory sequence is a monocot embyro-
25 specific promoter.

        14.   The chimeric gene of Claim 1, wherein the nucleic acid fragment encoding dihydrodipicolinic acid synthase comprises the nucleotide sequence shown in SEQ ID NO:3: encoding dihydrodipicolinic acid synthase from
30 Corynebacterium glutamicum and wherein the plant chloroplast transit sequence is derived from a gene encoding the small subunit of ribulose 1,5-bisphosphate carboxylase from Zea maize, and the seed-specific regulatory sequence is from the globulin 1 gene from Zea
35 maize.

15.   A method for obtaining a monocot plant wherein
the seeds of the plant accumulate lysine at a level from
ten percent to one hundred thirty percent higher than do
seeds of an untransformed plant comprising:

5   (a)   transforming monocot cells with the
chimeric gene of Claim 13 or 14;

           (b)   regenerating fertile mature plants from
the transformed plant cells obtained from step (a) under
conditions suitable to obtain seeds;

10   (c)   screening the progeny seed of step (b)
for lysine content; and

           (d)   selecting those lines whose seeds contain
increased levels of lysine.

16.   A method for obtaining a corn plant wherein
15   the seeds of the plant accumulate lysine at a level from
ten percent to one hundred thirty percent higher than do
seeds of an untransformed plant comprising:

           (a)   transforming corn cells with the chimeric
gene of Claim 13 or 14;

20   (b)   regenerating fertile mature plants from
the transformed plant cells obtained from step (a) under
conditions suitable to obtain seeds;

           (c)   screening the progeny seed of step (b)
for lysine content; and

25   (d)   selecting those lines whose seeds contain
increased levels of lysine.

17.   A monocot plant comprising in its genome the
chimeric gene of Claim 13 or 14.

18.   Seeds obtained from the plant of Claim 17.

30   19.   A transformed monocot plant wherein the seeds
of the plant accumulate lysine to a level between
thirty-five percent and one hundred thirty percent
higher than do seeds of an untransformed plant.

20.   A transformed monocot plant of Claim 19
35   wherein the plant is corn.

21. A nucleic acid fragment comprising:

(a)   a first chimeric gene of Claim 1, 2, 13 or 14 and

(b)   a second chimeric gene wherein a nucleic
5   acid fragment encoding a lysine-rich protein, wherein
the weight percent lysine is at least 15%, is operably
linked to a plant seed-specific regulatory sequence.

22. A nucleic acid fragment comprising:

(a)   a first chimeric gene of Claim 1, 2, 13
10   or 14 and

(b)   a second chimeric gene wherein a nucleic
acid fragment encoding a lysine-rich protein comprises a
nucleic acid sequence encoding a protein comprising n
heptad units (d e f g a b c), each heptad being either
15   the same or different, wherein:

n is at least 4;

a and d are independently selected from
the group consisting of Met, Leu,
Val, Ile and Thr;

20   e and g are independently selected from
the group consisting of the acid/base
pairs Glu/Lys, Lys/Glu, Arg/Glu,
Arg/Asp, Lys/Asp, Glu/Arg, Asp/Arg
and Asp/Lys; and

25   b, c and f are independently any amino
acids except Gly or Pro and at least
two amino acids of b, c and f in each
heptad are selected from the group
consisting of Glu, Lys, Asp, Arg,
30   His, Thr, Ser, Asn, Ala, Gln and Cys,
said nucleic acid fragment is operably linked to a plant
seed-specific regulatory sequence.

23. A nucleic acid fragment comprising:

(a)   a first chimeric gene of Claim 1, 2, 13
35   or 14 and

162

(b)   a second chimeric gene wherein a nucleic
acid fragment encoding a lysine-rich protein comprises a
nucleic acid sequence encoding a protein having the
amino acid sequence $(MEEKLKA)_6(MEEKMKA)_2$ is operably
linked to a plant seed-specific regulatory sequence.

24.   A plant comprising in its genome the chimeric
gene of Claim 1, 2, 13 or 14 and the second chimeric
gene of Claim 21, Claim 22 or Claim 23.

25.   A plant comprising in its genome the nucleic
acid fragment of Claim 21, Claim 22 or Claim 23.

26.   Seed obtained from the plant of Claim 24.

27.   Seed obtained from the plant of Claim 25.

28.   A nucleic acid fragment comprising

(a)   a first chimeric gene of Claim 1, 2, 13 or
14 and

(b)   a second chimeric gene wherein a nucleic
acid fragment encoding a lysine ketoglutarate reductase
is operably linked in the sense or antisense orientation
to a plant seed-specific regulatory sequence.

29.   A plant comprising in its genome the first
chimeric gene of Claim 1, 2, 13 or 14 and a second
chimeric gene wherein a nucleic acid fragment encoding a
lysine ketoglutarate reductase is operably linked in the
sense or antisense orientation to a plant seed-specific
regulatory sequence.

30.   A plant comprising in its genome the nucleic
acid fragment of Claim 28.

31.   Seed obtained from the plant of Claim 30.

FIG.1
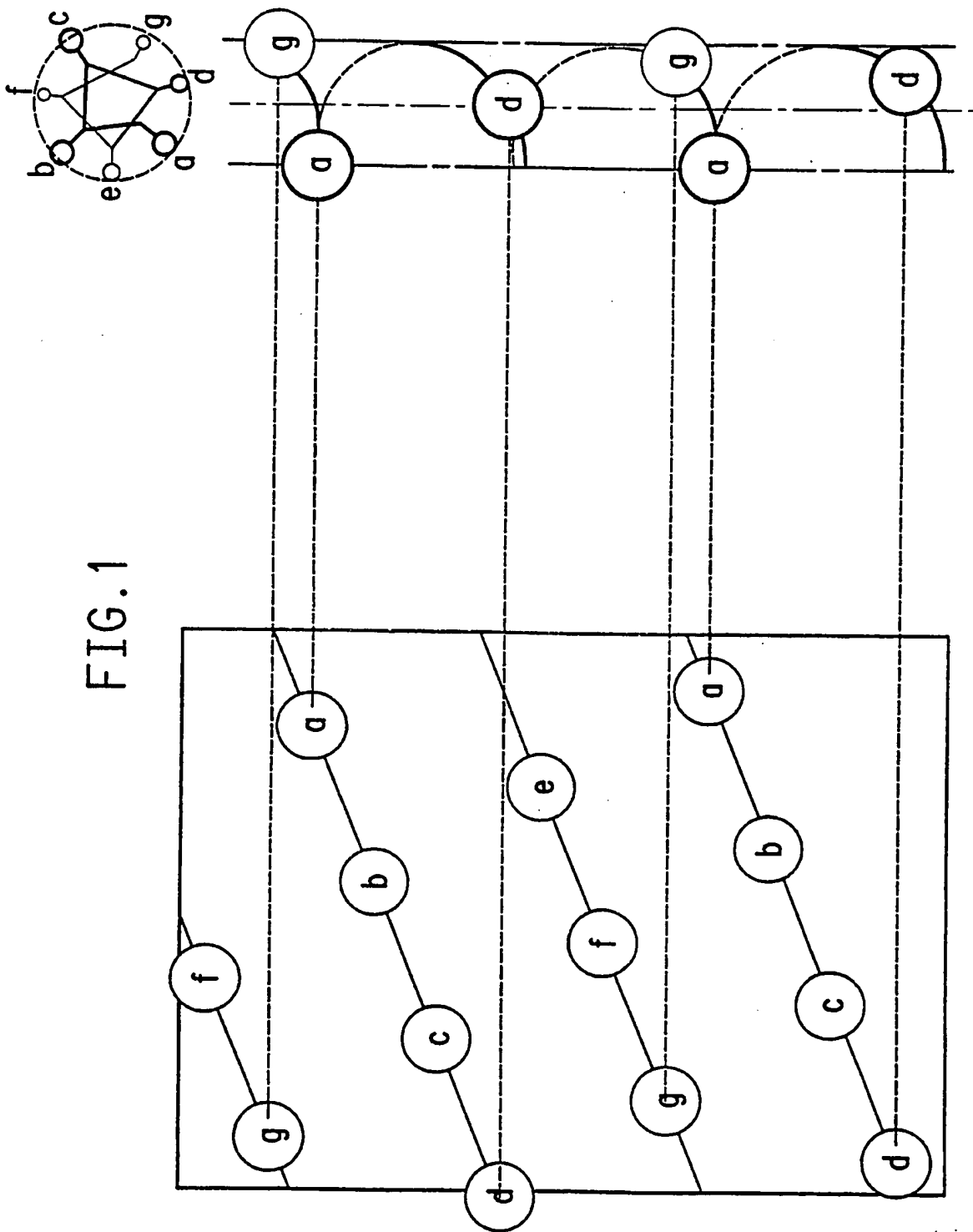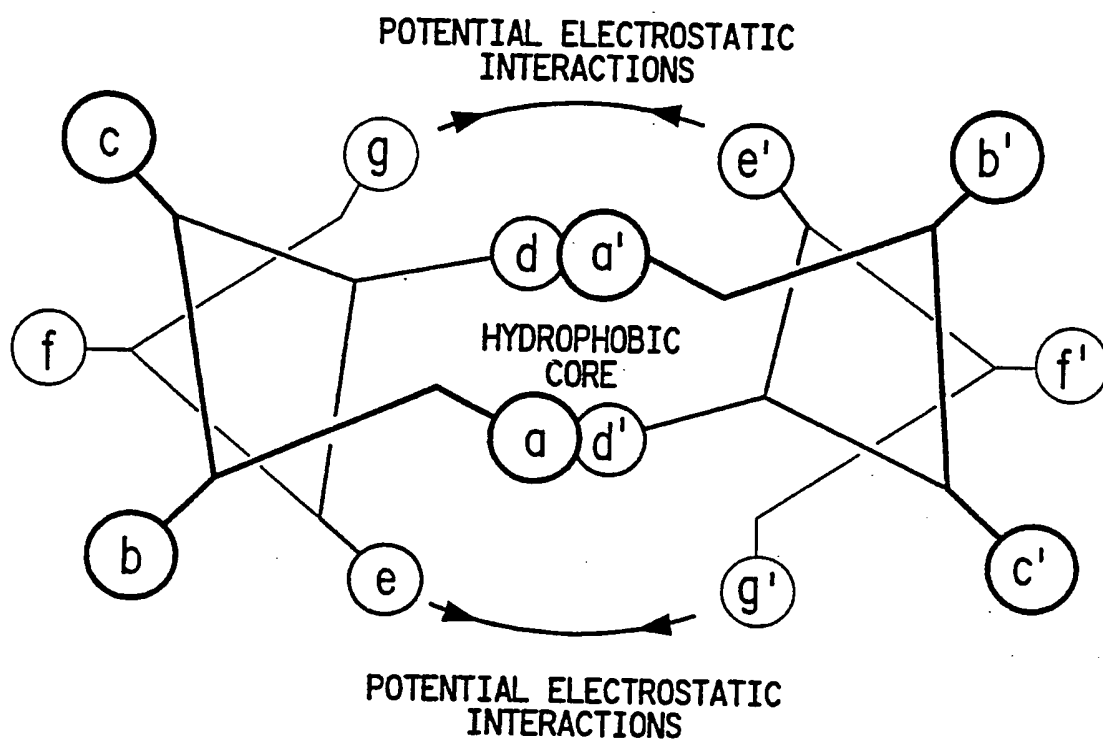
FIG.2A

POTENTIAL ELECTROSTATIC
INTERACTIONS

HYDROPHOBIC
CORE

POTENTIAL ELECTROSTATIC
INTERACTIONS

FIG.2B

FIG.3A

LEUCINE

FIG.3B

METHIONINE

FIG.4

HindIII

NcoI
SmaI
KpnI
XbaI

HindIII

500bp

1180bp

5'-phaseolln

3'-phaseolln

FIG.5A

FIG.5B

FIG.6A

FIG.6B

FIG.7

FIG.8

ASP718ECORI

EARI

NCOI
CATGGAGGAGAAGATGAAGGC GATGGAGGAGAAGATGAAGGCGTGATAGGTACCG
CTCCTCTTCTACT TCCGCTA CCTTCTC TTCGACTTCCGCACTATCCATGGCTTAA
M E E K M K A M E E K M K A
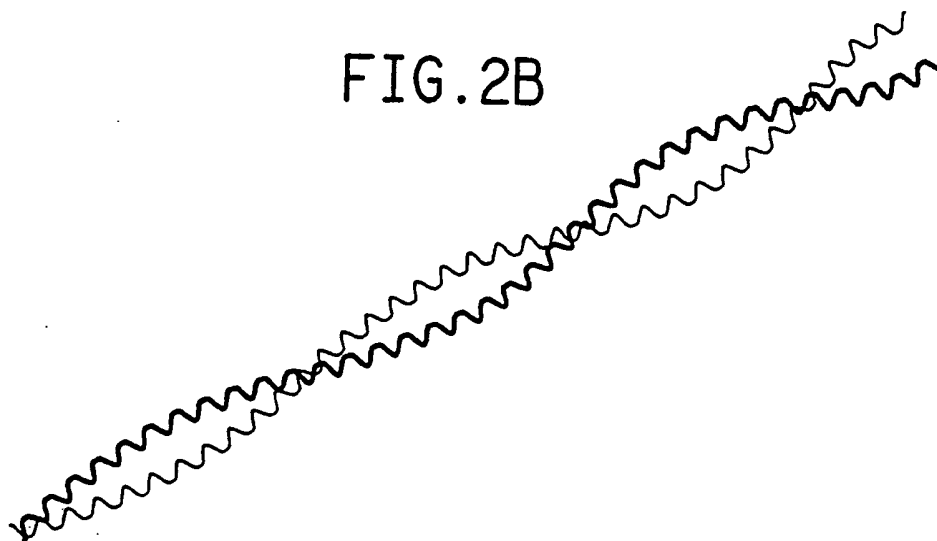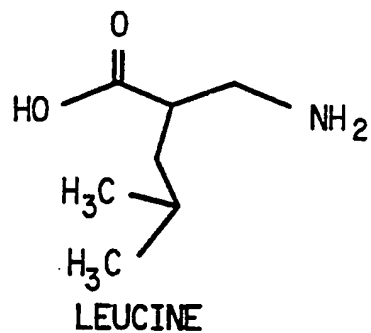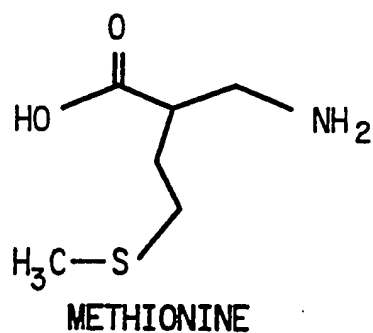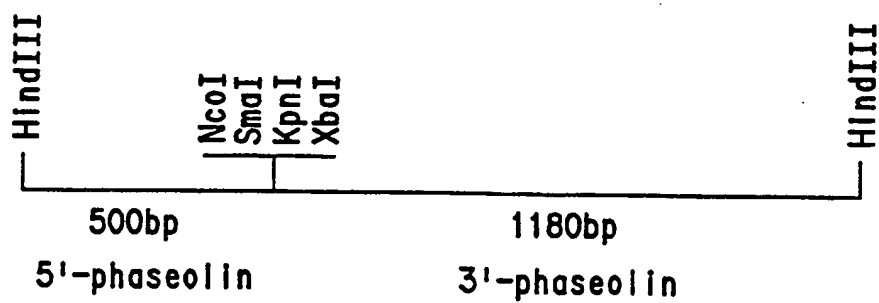
↓ LIGATE OLIGOS
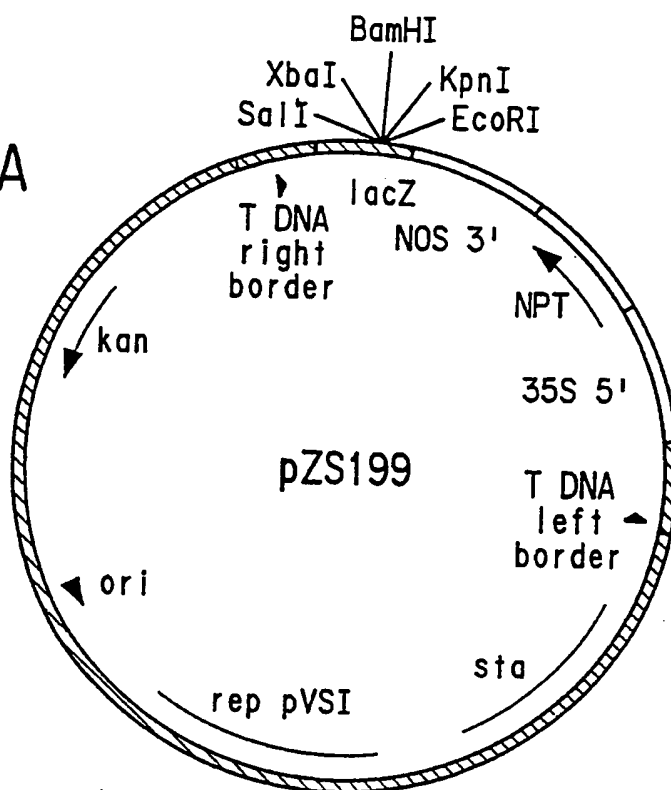
GATGGAGGAGAAGATGAAGGC
CCTCCTCTTCTACTTCCGCTA
M E E K M K A
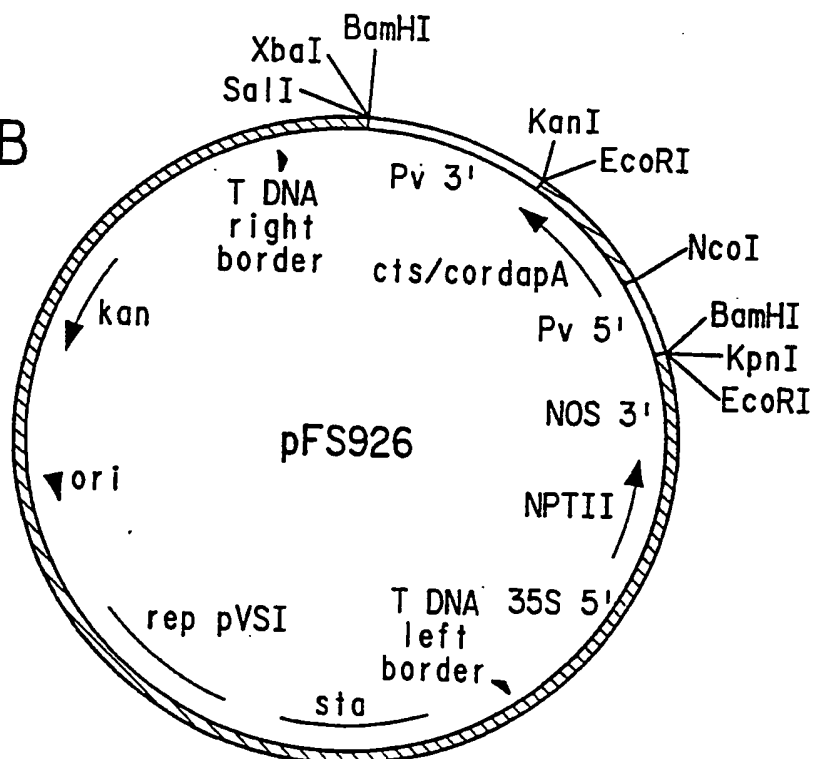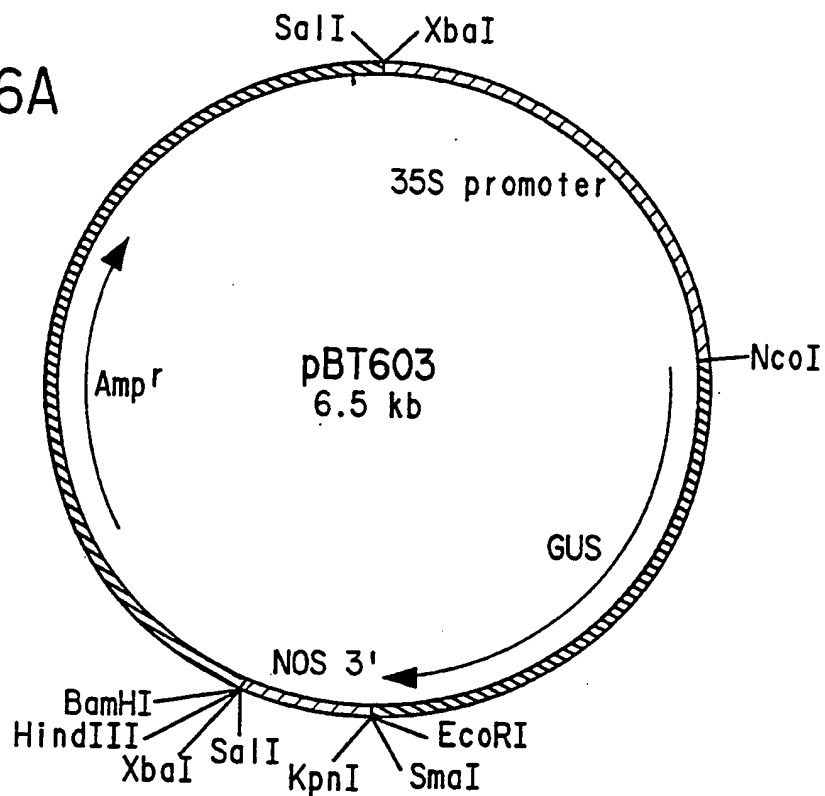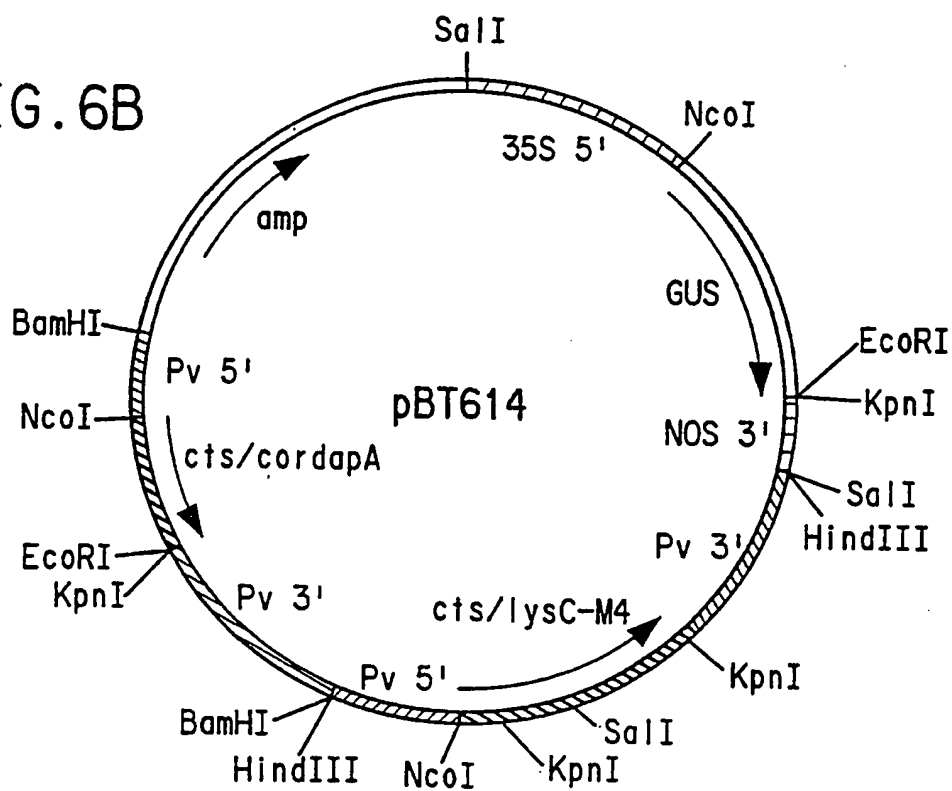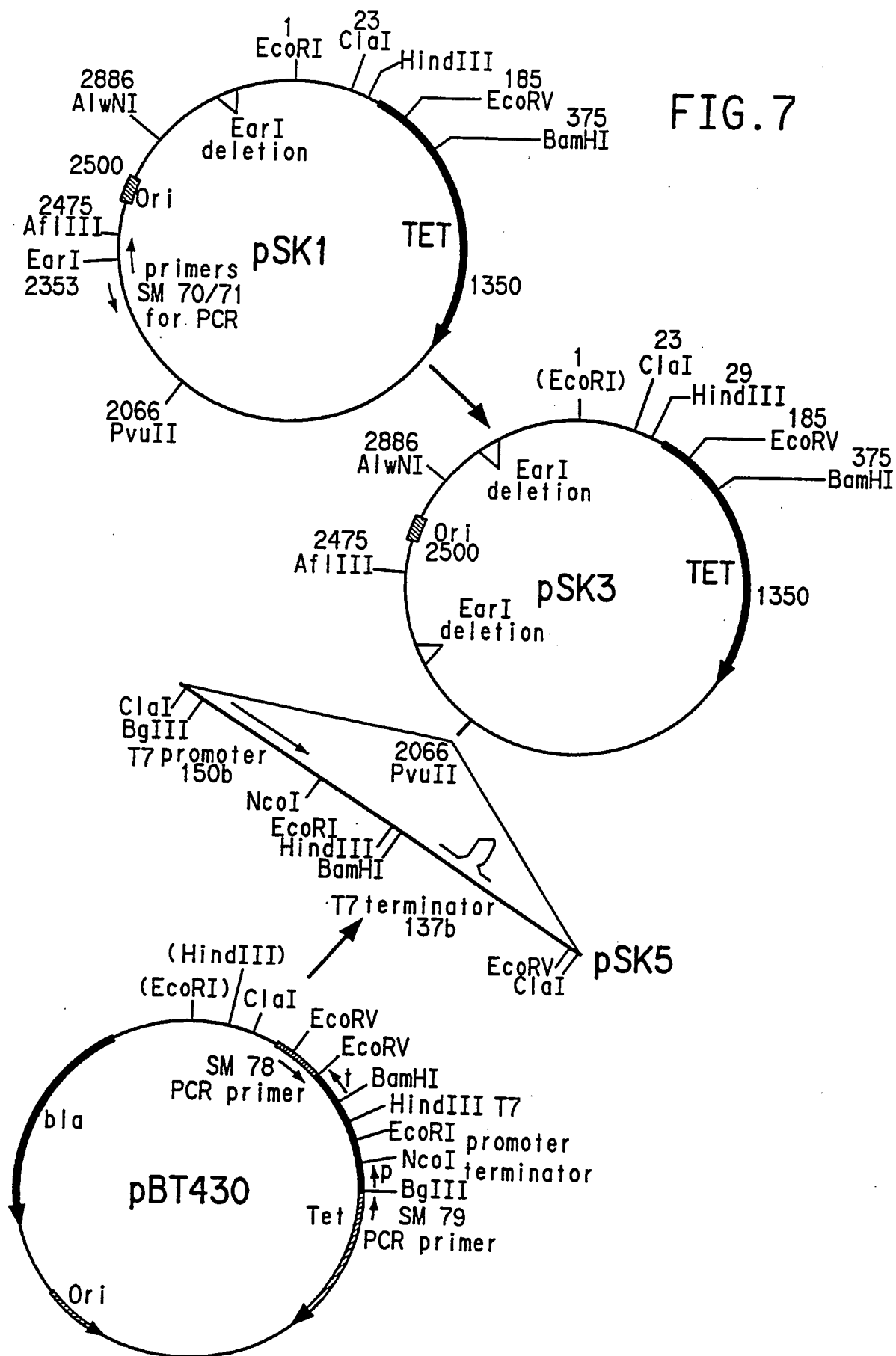
↓ LIGATE TO EARI CUT VECTOR

EARI   ASP718ECORI

NCOI
CATGGAGGAGAAGATGAAGGC GATGGAGGAGAAGATGAAGGCGTGATAGGTACCG
CTCCTCTTCTACT TCCGCTA CCTCCTCTTCTACTTCCGCTA CCTTCTC TTCGACTTCCGCACTATCCATGGCTTAA
M E E K M K A M E E K M K A M E E K M K A

FIG.9

FIG. 10

NCOI   EARI   BSPHI   STOP   ASP718   ECORI   BASE GENE

```
[CATGG]AGGAGAAGATGAAAAA      GCTC[GAAGAG]AAGATGAAGG[TCATGAAG]TGAT[AGGTACC]G
    CTCCTCTTCTACTTTTTCTA  GCTTCTCTTCTACTTCCCAGTACTTCACTATCCATGGCTTAA
M   E  E  K  M  K  K           L  E  E  K  M  K  K  V  M  K
```
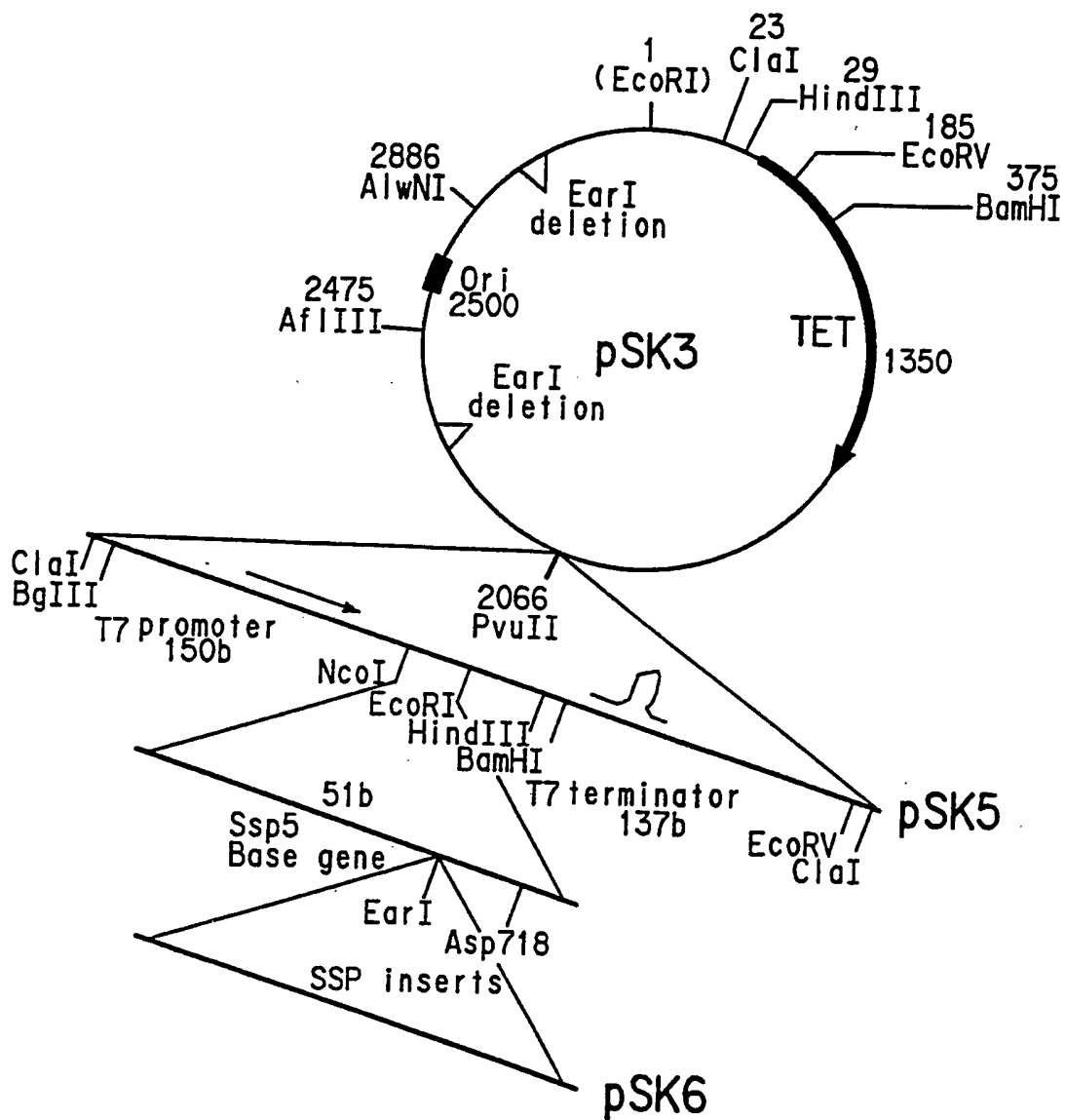
OLIGONUCLEOTIDE INSERTS

```
GCTGGAGAAAAGATGAAGGCTATGGAGGAGAAGATGAAGAA
CCTTCTTTTTCTACTTCCGATACCTCCTCTTCTACTTCTTCGA
 L  E  E  K  M  K  A  M  E  E  K  M  K  K
```

↓ OLIGOS LIGATED INTO EARI CUT BASE GENE

```
[CATGG]AGGAGAAGATGAAAAA GCTGGAAGAGAAGATGAAGGCTATGGAGGAGAAGATGAAGGCTTGAGGAAAGATGAAGAAGCT
    CTCCTCTTCTACTTTTTCTA CCTTCTTTTTCTACTTCCGATACCTCCTGTTCTACTTTACCGAACTCCTTTCTACTTCTTCGA
M   E  E  K  M  K  K  L  E  E  K  M  K  A  M  E  E  K  M  K  W  L  E  E  K  M  K  K  L
```

NCOI

```
[CATGG]AGGAGAAGATGAAAAA
    CTCCTCTTCTACTTTTTCTA
M   E  E  K  M  K  K
```

EARI   BSPHI   ASP718   ECORI   CLONE pSK34

```
C[GAAGAG]AAGATGAAGG[TCATGAAG]TGAT[AGGTACC]G
GCTTCTCTTCTACTTCCCAGTACTTCACTATCCATGGCTTAA
  E  E  K  M  K  K  V  M  K
```

FIG.11A

NheI (260)

BspHI (520)

ori

tet^r

Segment 3
or
Segment 4
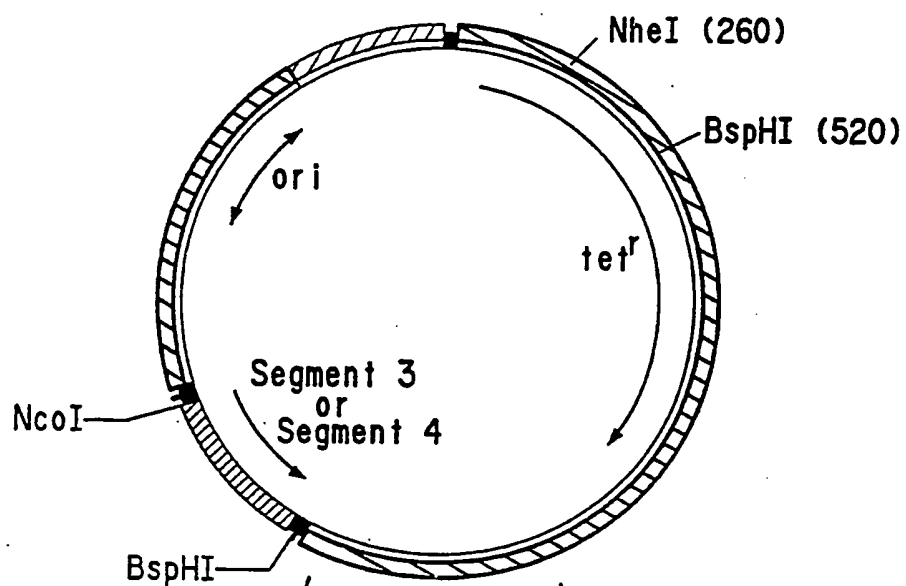
NcoI

BspHI

Segment 3
Digest NheI/BspHI

Segment 4
Digest NheI/NcoI

FIG.11B

NheI   NheI

BspHI (520)

ori

LIGATE FRAGMENTS

tet^r

Segment 3

BspHI

Segment 4

NcoI
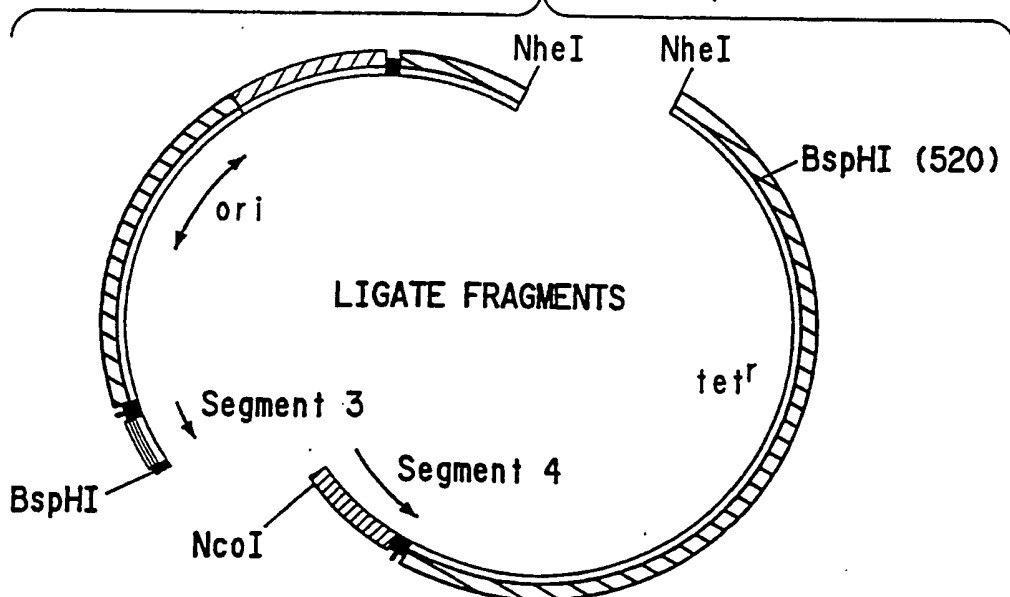
FIG.12
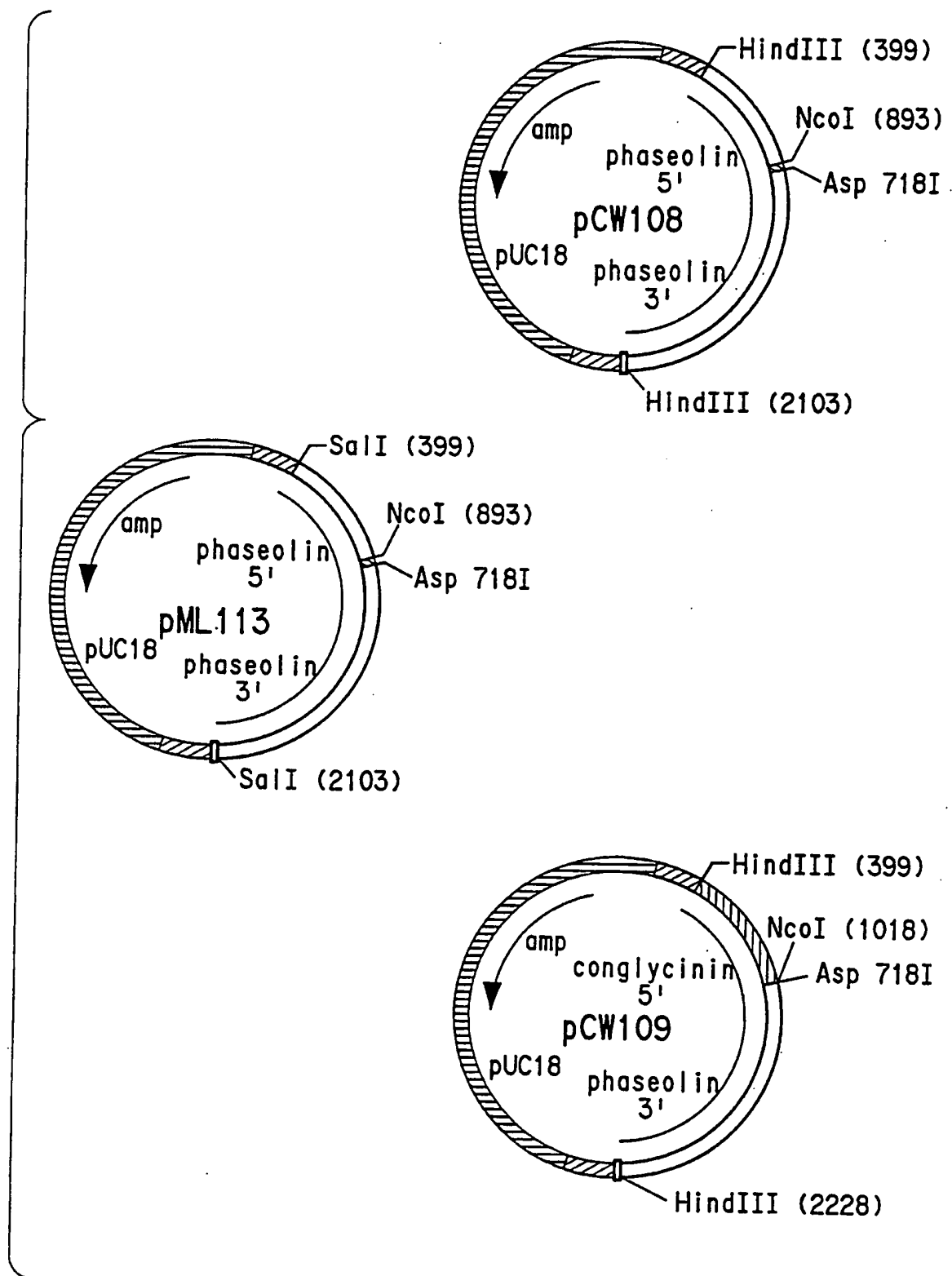
FIG.13
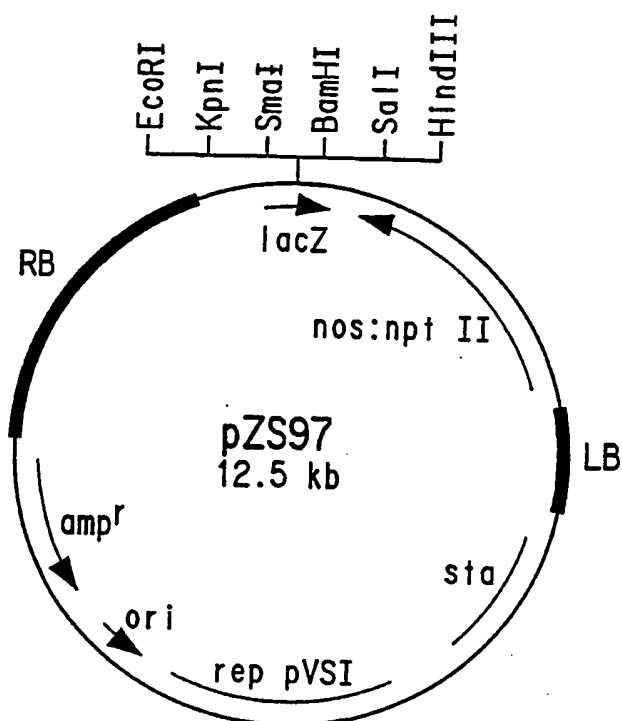
FIG.14
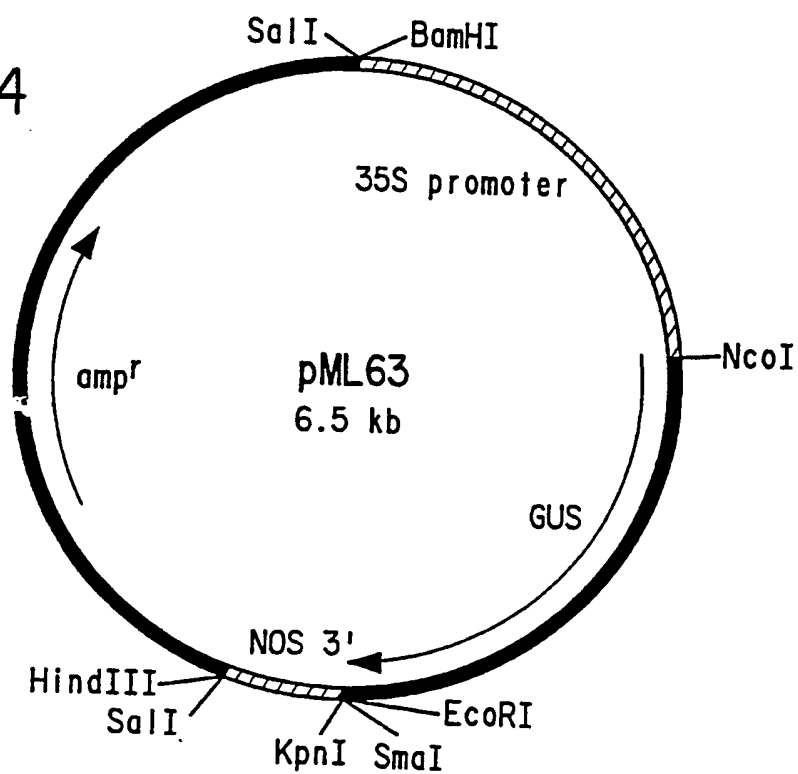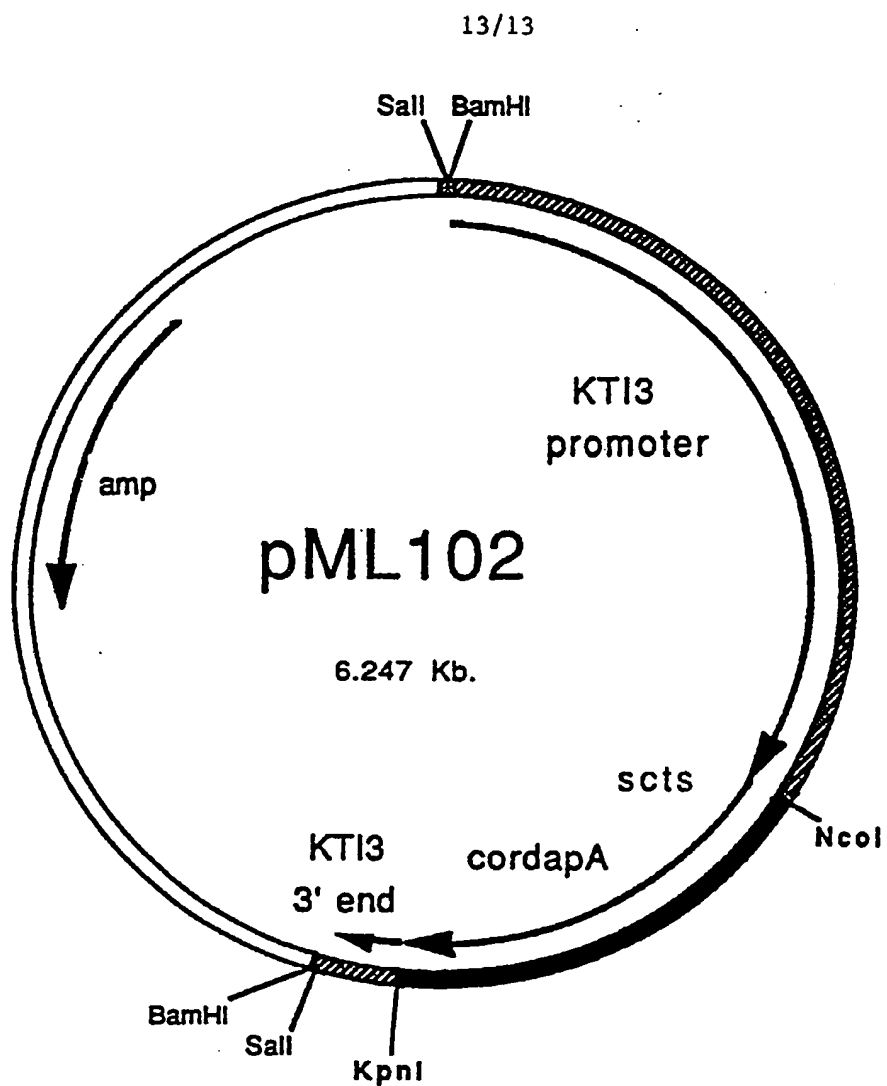
FIGURE 15

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6   C12N15/82   C12N15/60   C12N15/11   A01H5/00   A01H5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6   C12N   A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,93 19190 (DU PONT) 30 September 1993 see examples 3,6,10,12,15 see examples 16,19 ---	1-6,9,10
P,X	PLANT PHYSIOLOGY SUPPLEMENT, vol.105, no.1, May 1994 page 115 SANDERS, C., ET AL. 'Overproduction of lysine in seeds of Brassica napus cv westar by genetic engineering of the biosynthetic pathway' see abstract 617 --- -/--	9-11

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 March 1995	

| Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016 | Authorized officer

Maddox, A |

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	JOURNAL OF CELLULAR BIOCHEMISTRY SUPPLEMENT, vol.18A, 1994 page 79 FALCO, S.C., ET AL. 'Transgenic crops with improved amino acid composition' see abstract X1-017 & KEYSTONE SYMPOSIUM ON IMPROVED CROP AND PLANT PRODUCTS THROUGH BIOTECHNOLOGY, KEYSTONE, COLORADO, USA, JANUARY 9-16, 1994.	1-5
P,A	JOURNAL OF CELLULAR BIOCHEMISTRY SUPPLEMENT, vol.18A, 1994 page 107 RICE, J.A., ET AL. 'Expression of synthetic high lysine seed storage proteins can significantly increase the accumulated levels of lysine in mature seeds of transgenic crop plants' see abstract X1-329 & KEYSTONE SYMPOSIUM ON IMPROVED CROP AND PLANT PRODUCTS THROUGH BIOTECHNOLOGY, KEYSTONE, COLORADO, USA, JANUARY 9-16, 1994.	21-31
A	WO,A,93 03160 (DU PONT) 18 February 1993 see the whole document	21-31

2

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9319190	30-09-93	AU-B- 3923393	21-10-93
		CA-A- 2132414	30-09-93
		EP-A- 0640141	01-03-95
WO-A-9303160	18-02-93	AU-A- 2441292	02-03-93
		CA-A- 2114788	18-02-93
		EP-A- 0598806	01-06-94